Contribution to the understanding of mental task BCI performances using predictive computational models

Camille Benaroch

université de BORDEAUX

université de BORDEAUX

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR
DE L'UNIVERSITÉ DE BORDEAUX

ECOLE DOCTORALE MATHEMATIQUES ET INFORMATIQUES

ECOLOGIE ÉVOLUTIVE, FONCTIONNELLE ET DES COMMUNAUTÉS

Par **Camille BENAROCH**

Contribution to the understanding of mental task BCI performances using predictive computational models

Sous la direction de : **Fabien LOTTE**
Encadrante : **Camille JEUNET**

Soutenance le 17 décembre 2021

Membres du jury :

Pr. Moritz GROSSE-WENTRUP	Professeur des université	Université de Vienne	Rapporteur
Dr. Nathalie GEORGE	Directrice de Recherche	Institut du Cerveau / CNRS	Rapporteur
Pr. François CABESTAING	Professeur des université	Université de Lille	Président
Dr. Maryam ALIMARDANI	Professeur assistant	Université de Tilburg	Examinatrice
Dr. Fabien LOTTE	Directeur de recherche	Inria Bordeaux Sud-Ouest	Directeur
Dr. Camille JEUNET	Chargée de Recherche	CNRS / Université de Bordeaux	Encadrante

Contribution à la compréhension des performances BCI basées sur les tâches mentales à l'aide de modèles computationnels prédictifs

Résumé : Les interfaces cerveau-ordinateur (ICO) sont des outils de communication et de contrôle qui permettent à leurs utilisateurs d'interagir avec un ordinateur via leur activité cérébrale (mesurée, généralement, à l'aide de l'électroencéphalographie - EEG). Une catégorie prometteuse d'ICO est l'ICO basée sur les tâches mentales (TM). Les TM-ICO utilisent les modifications de l'activité cérébrale induites par les TM effectuées par l'utilisateur (par exemple, l'imagination de mouvements, le calcul mental ou la rotation mentale d'un objet) pour les transformer en commandes de contrôle. Contrôler une TM-ICO nécessite l'acquisition de compétences et donc un entraînement approprié. En effet, l'utilisateur doit générer des signaux cérébraux stables et distincts pour chaque tâche, faute de quoi il ne sera pas en mesure de contrôler le système. En effet, le système ne sera pas en mesure de reconnaître quelle tâche l'utilisateur est en train d'effectuer. Produire de tels signaux cérébraux est une compétence à acquérir et à maîtriser. L'objectif de cette thèse est de contribuer à la compréhension de l'entraînement des utilisateurs d'ICO en réalisant d'abord une étude expérimentale de l'apprentissage. Dans une première partie, nous avons proposé et évalué la conception d'une TM-ICO multi-classes pour entrainer un utilisateur tétraplégique sur le long terme. Nou avons utilisé une nouvelle méthode de classification: les classifieurs riemanniens adaptatifs. Nous avons également observé que notre pilote a appris à améliorer l'ICO en produisant des signaux EEG correspondant de plus en plus à la distribution des données d'entraînement du classificateur, plutôt qu'en améliorant à discriminer ses signaux. Cette étude nous a également permis de constater la difficulté de la mise en place d'un protocole fiable dédié à un entraînement ICO à long terme. La seconde partie de notre travail est consacrée à la compréhension des performances des TM-ICO à l'aide de modèles computationnels prédictifs. Nous avons proposé différents modèles pouvant prédire les performances de différents utilisateurs de ICO au cours de l'entraînement basés sur des caractéristiques liées aux ICO. Comme une ICO est un système de communication entre un utilisateur et une machine, ces caractéristiques sont liés à la fois au profil de l'utilisateur at aux facteurs extraits d'algorithmes utilisés pour construire/calibrer le système. Nos résultats suggèrent qu'il est possible de prédire les performances des utilisateurs d'ICO en utilisant les caractéristiques neurophysiologiques d'un utilisateur, mais aussi les caractéristiques neurophysiologiques combinées à des caractéristiques stables (des traits) de l'utilisateur. De plus, nos études ont révélé que l'étude des caractéristiques extraites des méthodes utilisées pour construire/calibrer le système pourraient être intéressantes pour mieux comprendre pourquoi certains sujets ont des difficultés à contrôler une ICO. En effet, des modèles fiables de performances ont été révélés en utilisant de telles caractéristiques.

Mots-clés : Interfaces Cerveau-Orinateur, Modélisation Computationnelle, Entraînement Utilisateur, Apprentissage, Profil, Neurophysiologie, Classification, Traitement de Données, ElectroEncéphaloGraphie

Contribution to the understanding of mental task BCI performances using predictive computational models

Abstract: Brain computer interfaces (BCIs) are communication and control tools that enable their users to interact with computer by using brain activity alone (which is measured, most of the time, using electroencephalography - EEG). A prominent type of BCI is mental task (MT) based BCIs, that translate modifications in brain activity induced by MTs performed by the user (e.g., imagination of movements, mental calculation or mental rotation of an object among others) into control commands for a computer. Using an MT-BCI requires dedicated training. Indeed, the user has to generate stable and distinct brain signals for each task otherwise they will not be able to control the system. Indeed, the system will not be able to recognize which task the user is performing. Producing such brain signals is a skill to be acquired and mastered and the more the user practices the better he/she will get at it. The objective of my PhD project is to contribute to the understanding of BCI user training by first doing an experimental study of learning by participating in the CYBATHLON competition. We proposed and evaluated the design of a multi-class MT-based BCI for longitudinal training of a tetraplegic user with a newly designed machine learning pipeline based on adaptive Riemannian classifiers. Using a newly proposed BCI user learning metric, we could show that our user learned to improve his BCI control by producing EEG signals matching increasingly more the BCI classifier training data distribution, rather than by improving his EEG class discrimination. In addition, this study revealed the difficulty of setting up a reliable protocol dedicated to a long term BCI training. The second part of this work is dedicated to the understanding of MT-BCI performances using predictive computational models. We proposed various computational models of BCI user training that could predict the performances of various BCI users over training time, based on BCI systems component. As a BCI is a communication system between a user and a machine such components were related to the user-profile related characteristics but also factors extracted from machine-learning algorithms used to build the system classifier. Our results suggested that is was possible to predict BCI performances using neurophysiological characteristics of a user but also neurophysiological characteristics combined with stable characteristics (i.e., traits) or the user. In addition, our studies revealed that studying features extracted from data-driven methods could be interesting to better understand why some subjects have difficulties controlling a BCI. Indeed, reliable models of BCI performances were revealed using such features.

Keywords: Brain-Computer Interface, Computational Modeling, User Training, Learning, Profile, Neurophysiology, Classification, Processing, ElectroEncephaloGraphy

Equipe POTIOC Inria/ LABRI
UMR 5800 Université de Bordeaux, 33000 Bordeaux, France.

ACRONYMS

BCI	Brain-Computer Interface
MT	Mental-Task
MI	Mental-Imagery
EEG	ElectroEncephalography
NFT	NeuroFeedback Training
SCP	Slow Cortical Potentials
ERP	Event-Related Potential
ERD	Event-Related Desynchronisation
ERS	Event-Related Synchronisation
SMR	SensoryMotor Rythm
CNS	Central Nervous System
CSP	Common Spatial Pattern
ICA	Independant Component Analysis
LDA	Linear Discriminant Analysis
SCM	Spatial Covariance Matrix
TTA	Test Train Adaptation
FgMDM	Fisher Geodesic Minimum Distance to Mean
FGDA	Fisher Geodesic Discriminant Analysis
MDBF	Most Discriminant Frequency Band
UX	User eXperience
LASSO	Least Absolute Shrinkage and Selection Operator
EN	Elastic-Net
16pf5	16 Personnality Factors - version 5
SD	Standard Deviation
EO	Eyes-Open
EC	Eyes-Closed
PSD	Power Spectral Density
BH	Benjamini-Hochberg

Remerciements

Après trois ans de travail pour cette thèse, un manuscript enfin finalisé et une soutenance passée, il est temps de remercier toutes les personnes qui ont rendu tout cela possible et sans qui rien n'aurait été possible. J'espère n'oublier personne.

Je voudrais commencer par remercier les deux personnes sans qui ce travail n'existerait pas, deux personnes que je ne pourrais jamais assez remercier tellement ils sont exceptionnels: Fabien et Camille. Merci pour tout! Merci de m'avoir fait confiance pour effectuer le stage qui a précédé cette thèse et merci de m'avoir proposé de continuer ce travail avec vous. Merci de m'avoir si bien encadrée, d'avoir toujours pris le temps de discuter, corriger et relire mon travail. Merci d'avoir toujours été disponibles, même à l'autre bout du monde. Je n'aurais pas pu ésperer mieux comme encadrants, vous êtes le meilleur duo pour ce job! Fabien, merci pour ta patience et ta bienveillance. Camille, je sais que j'ai eu l'immense privilège d'être la première personne que tu encadrais pour une thèse et je voudrais te dire que tu as été parfaite et que les prochains doctorants que tu encadreras auront BEAUCOUP de chance.

Si j'ai passé trois belles années de thèse c'est aussi grâce à toute l'équipe POTIOC. Il y a eu ceux du début de ma thèse, Pierre-Antoine, Damien, Thibault, Sol, Lauren, Jelena, Léa, Cécile, Théo et Erwan, mais aussi ceux qui sont arrivés pendant le COVID et à la fin de ma thèse, Arnaud, Sebastien, Pauline, Ambre, Edwige, Moargane, David, Luiz, Pierre et Yvonne et enfin ceux qui ont toujours (ou presque) été à mes cotés Audrey, Martin, Pascal, Aline, Raj, khadijeh, Adé, Thibaut et surtout les Mac lovers, Aurélien et Philippe, mon bureau des plaintes, mes grands frères de thèse, qui m'ont montré que même si il y avait des moments difficiles, on arrivait toujours au bout. Enfin, merci à tous les étudiants avec qui j'ai eu la chance de travailler, Eidan, Maria et Alper.

Faire une thèse c'est aussi aller à l'étranger, faire des conférences et rencontrer du monde. Malheureusement, dans mon cas le covid en aura décidé autrement. Cependant, j'ai quand même eu la chance de vivre une expérience unique pendant le cybathlon à Graz. Alors merci à toute l'équipe mais surtout à Fabien et Wilfried qui ont rendu cela possible.

Je voudrais aussi remercier toutes les personnes qui ont assisté à ma soutenance, sur place ou en ligne. Votre présence à tous m'a vraiment fait chaud au coeur! Et plus particulièrement, je voudrais remercier mon jury, qui m'a fait l'honneur de bien vouloir relire et évaluer ce manuscrit. Merci Nathalie George, François Cabestaing. Thank you Moritz Grosse-Wentrup and Maryam Alimardani. Thank you all for your precious advice and your kindness. Merci aussi à ma famille pour leur soutien tout au long de ce doctorat et plus particulièrement pendant les mois de télétravail! Merci à mes parents pour leur confiance et leur amour inconditionnels, pour leur soutien de tous les instants et pour leurs encouragements. Je remercie aussi mes sœurs Pauline et Noémie et mon frère Simon pour leurs soutiens et attentions. Ils m'ont permis de réaliser que la famille est sacrée. Ils étaient pour moi, une vraie source d'inspiration et ont été toujours à mes côtés durant les moments difficiles. Merci Anne-Marie de m'avoir accueillie dans ta famille en me considérant comme un fille. Merci Gyslaine et Bernard de m'avoir acceuilli chez vous au début de ma thèse.

Pour finir, je remercie Michel qui m'a toujours soutenue malgrès les 587 km qui nous séparaient au quotidien. Je t'aime.

Resumé en français

Introduction

Les Interfaces Cerveau-Ordinateur (ou BCI de l'anglais *Brain-Computer Interface*) sont des systèmes de communication et contrôle permettant à un utilisateur d'envoyer des commandes à un ordinateur sans aucune activité physique [1, 2]. Typiquement, la BCI mesure et reconnait l'activité cérébrale (généralement par électroencéphalographie ou EEG) associée à l'intention de l'utilisateur et la convertit en une commande pour une application. Par exemple, une BCI peut reconnaître dans les signaux EEG d'un utilisateur qu'il imagine un mouvement de la main gauche ou droite, pour faire un bouger un curseur à l'écran vers la gauche ou droite, respectivement. Puisque les BCI permettent une interaction directe entre le cerveau et le monde extérieur, elles promettent de révolutionner l'interaction humain-machine en général, par exemple pour les technologies d'assistance ou les jeux vidéo. Malgré leurs promesses, les BCI sont encore très peu utilisées hors des laboratoires, notamment à cause de leur mauvaise fiabilité. Par exemple, les BCI actuelles utilisant deux mouvements imaginés des mains (on parle de Tâche Mentale - TM) comme commandes mentales, décodent correctement moins de 80% de ces commandes en moyenne [3], tandis que 10 à 30% des utilisateurs ne parviennent pas du tout à contrôler ces BCI [4]. Ainsi, pour que les BCI puissent véritablement être utiles pour des personnes gravement paralysées et des utilisateurs en bonne santé, il faut qu'elles deviennent fiables et utilisables en pratique, hors des laboratoires. Jusqu'à présent, la majorité des recherches en BCI ont été dédiées au traitement et à la classification des signaux EEG, avec la conception et l'étude d'une pléthore d'algorithmes d'apprentissage artificiel (ou "Machine Learning" - ML) [5]. Si ces recherches ont permis des progrès, elles ont cependant trop souvent ignoré un élément essentiel d'une BCI : l'utilisateur. En effet, contrôler une BCI est une compétence qui s'apprend et qu'il faut entraîner, notamment pour les BCI utilisant des TM (TM-BCI) [1]. Si l'utilisateur ne peut pas produire des motifs EEG stables et distincts, alors peu importe l'algorithme d'analyse des signaux EEG utilisé, ce dernier ne pourra pas reconnaître ces motifs EEG. Ainsi, une BCI est un système de communication co-adaptatif: l'apprentissage de la machine doit être couplé à des sessions d'entraînement de l'utilisateur afin d'atteindre une performance de contrôle acceptable.

Les résultats de recherches récentes ont montré que la façon dont les utilisateurs de BCI sont actuellement entraînés est sous-optimale, tant sur le plan théorique [6, 7] que pratique [8]. L'une des raisons pourrait être que chaque utilisateur commence avec des compétences, une activité cérébrale et des traits de personnalité différents, mais avec le même protocole d'entraînement. De plus, l'utilisateur est connu pour être l'une des principales causes de la variabilité des signaux EEG dans les BCI, en raison de ses changements d'humeur, de sa fatigue, de son attention, etc. [9]. Par conséquent, l'adaptation du protocol d'entraînement à chaque utilisateur est un domaine de recherche prometteur pour améliorer la fiabilité des BCI. C'est dans ce contexte que nous considérons comme essentiel de comprendre l'apprentissage des MT-BCI et d'identifier les facteurs qui l'influencent. En effet, pour améliorer et adapter l'apprentissage des utilisateurs, il est nécessaire de comprendre les facteurs ayant un impact sur le processus d'apprentissage des utilisateurs de TM-BCI et d'améliorer les protocoles d'entraînement en conséquence.

Par conséquent, parmi les nombreuses améliorations possibles qui peuvent être apportées à la recherche de prédicteurs des performances d'utilisateurs de TM-BCI pour adapter leur formation, nous allons aborder trois points principaux dans cette thèse : 1) la conception et l'étude d'un TM-BCI multi-classes pour l'entraînement sur le long terme (20 sessions sur 3 mois) d'un utilisateur final (c'est-à-dire, tétraplégique), 2) la création de modèles mathématiques pouvant prédire les performances de divers utilisateurs de BCI, au cours de leur entraînement, en fonction de leurs traits, de leurs compétences, de leurs caractéristiques neurophysiologiques mais aussi des caractéristiques de la machine, 3) l'étude de la généralisation des modèles prédictifs à travers différentes expériences et ensembles de données.

Entraînement à long terme d'un utilisateur tétraplégique par BCI

Bien que prometteuses, les TM-BCI ne sont généralement pas encore assez fiables pour être utilisées en pratique, c'est-à-dire dans des applications du *monde réel*, en dehors des laboratoires. En effet, il semble que le décodage des commandes mentales utiisées par les utilisateurs soit sujet à des taux d'erreur élevés [3, 4]. Il est donc nécessaire de concevoir des TM-BCI fiables à long terme, qui peuvent être utilisées en dehors du laboratoire par les utilisateurs finaux, par exemple les personnes souffrant d'un handicap moteur grave. Par conséquent, dans le chapitre Chapter 3, nous proposons et évaluons la conception d'un TM-BCI multi-classes pour l'entraînement longitudinal (20 sessions sur 3 mois) d'un utilisateur tétraplégique pour le CYBATHLON BCI series 2019. Dans ce championnat BCI, les pilotes tétraplégiques conduisent une voiture virtuelle en utilisant uniquement leur activité cérébrale, dans un jeu vidéo de course. Nous avons cherché à combiner l'entraînement d'un TM-BCI progressif de l'utilisateur avec une pipeline d'apprentissage automatique nouvellement conçue, basée sur des classifieurs riemanniens adaptatifs qui se sont révélés prometteurs pour les applications réelles. Nous avons suivi un processus de formation en deux étapes : les 11 premières sessions ont servi à entraîner l'utilisateur à contrôler un TM-BCI de 2 classes en effectuant soit deux tâches cognitives (REPOS et SOUSTRACTION MENTALE), soit deux tâches d'imagerie motrice (MAIN GAUCHE et MAIN DROITE). La deuxième étape de formation (9 sessions restantes) a appliqué un classificateur Riemannien adaptatif et indépendant de la session qui a combiné les 4 TM utilisées auparavant. Ce classificateur a pour but de traiter la variabilité entre les sessions. En effet, les classifieurs riemanniens atteignent généralement une précision élevée, ce qui rend prometteuses les applications BCI dans la vie réelle. De plus, comme notre classifieurs riemannien a été mis à jour de manière incrémentielle et non supervisée, il a pu capturer la non-stationnarité à l'intérieur et entre les sessions. Il est important de noter que nous rendons également compte de l'évolution des schémas neurophysiologiques et de l'expérience de l'utilisateur tout au long de la formation et de la compétition BCI afin de mieux comprendre l'apprentissage de l'utilisateur.

Nos résultats ont montré un apprentissage à tous les niveaux (c'est-à-dire, utilisateur, machine, et expérimentateurs). En effet, pendant les quelques mois d'entraînement, nous avons pu observer un apprentissage de l'utilisateur. La précision de la classification a augmenté de manière significative au cours de la formation, tandis que la métrique que nous avons proposée, c'est-à-dire l'adaptation test-train (TTA) reflétant le degré de distinction et de stabilité des tracés EEG produits par l'utilisateur indépendamment de tout classificateur, a diminué de manière significative, ce qui reflète l'apprentissage de l'utilisateur. En outre, nous avons pu proposer une nouvelle approche adaptative riemannienne pour réduire les décalages de la distribution EEG au sein d'une même session et entre les sessions. Cette approche pourrait, à l'avenir, être utilisée pour améliorer l'apprentissage de l'utilisateur en stabilisant le retour d'information du BCI. Comme le contrôle d'un BCI peut être long et difficile car la génération d'un signal cérébral distinct que le BCI peut reconnaître est une compétence à apprendre, nous pensons que l'amélioration de la technologie pour aider l'utilisateur à atteindre cet objectif est l'un des domaines de recherche sur lesquels nous devons nous concentrer. En plus de l'amélioration de la technologie, notre étude a montré qu'il est également essentiel de se concentrer sur la formation et l'expérience de l'utilisateur. Enfin, nous avons également beaucoup appris en faisant des erreurs pendant la formation. Cela nous a permis d'identifier plusieurs directions de recherche intéressantes pour l'avenir.

Cette expérience a révélé l'importance de pouvoir prédire les performances de l'utilisateur d'un TM-BCI afin de mettre en place le protocole de formation spécifique au sujet le plus optimal. En effet, au cours de cette étude, nous avons avancé à l'aveugle, en faisant des hypothèses et en modifiant l'entraînement assez souvent. Une première phase consacrée 1) à l'extraction des caractéristiques utilisées dans les modèles de performances TM-BCI et 2) à la prédiction des futures performances TM-BCI à l'aide de ces modèles aurait pu nous aider à concevoir un protocole d'entraînement pertinent. Ce processus aurait pu affecter plusieurs paramètres : la motivation des utilisateurs et des membres de l'équipe de recherche, les performances et l'organisation générale de la phase de formation. Les

parties suivantes de ce manuscrit sont donc consacrées à la compréhension des performances des utlisateurs de TM-BCI et à l'identification des facteurs qui les influencent en utilisant des modèles prédictifs des performances à travers les données issues de différentes expériences.

Les caracteristiques issues de l'utilisateur

Comme évoqué précédemment, le contrôle des TM-BCI est une compétence qui doit être acquise et maîtrisée par l'utilisateur [10–12]. En effet, l'efficacité des TM-BCI dépend intrinsèquement de la capacité des utilisateurs à produire des signaux EEG qui sont stables chaque fois qu'ils ont l'intention d'envoyer une même commande, et distincts entre les différentes commandes mentales. Pour rendre les TM-BCI plus fiables, les chercheurs se sont principalement concentrés sur les améliorations matérielles (par exemple, les électrodes) et logicielles (par exemple, les algorithmes de traitement du signal) mais moins sur l'amélioration des protocoles d'entraînement des utilisateurs. Pourtant, comme expliqué précédemment, cet aspect est également essentiel. En effet, les individus varient dans leurs caractéristiques personnelles à travers de nombreuses dimensions. Ces caractéristiques comprennent les traits démographiques, psychologiques, cognitifs et physiologiques de la personne. Chaque utilisateur ayant des caractéristiques différentes, il est nécessaire de développer des plans d'entraînement qui soient spécifiquement adaptées à chacun d'entre eux. Ce n'est pas le cas actuellement : les protocoles d'entraînement sont la plupart du temps génériques [6]. Afin d'approfondir notre compréhension des mécanismes sous-jacents au contrôle des TM-BCI, et par conséquent de concevoir des stratégies d'entraînement adaptées à chaque utilisateur, plusieurs études se sont intéressées aux prédicteurs de performance des TM-BCI [13–15]. Ces prédicteurs pourraient expliquer une partie de la variabilité entre les sujets en termes de capacités de contrôle de la TM-BCI. Cependant, pour l'instant, si un utilisateur débutant était incapable d'exécuter les commandes mentales souhaitées en raison de son incapacité à produire des schémas EEG appropriés, quel que soit le traitement du signal utilisé, nous pourrions avoir des difficultés à l'expliquer. L'identification des facteurs prédictifs de la performance des utilisateurs de TM-BCI pourrait nous aider à améliorer et à adapter le protocole de formation de l'utilisateur à chaque utilisateur.

Une catégorie de facteurs qui peuvent être intéressants pour prédire et expliquer les performances des utilisateurs de TM-BCI sont ceux basés sur des caractéristiques stables des utilisateurs. De tels facteurs peuvent être issus de leur profil de personnalité ou leur âge. Ces prédicteurs potentiels sont accessibles par le biais de questionnaires ou de tests. De plus, comme les BCI sont des neurotechnologies utilisant l'activité cérébrale de l'utilisateur pour contrôler des systèmes, pour mieux comprendre pourquoi certains utilisateurs parviennent à produire des signaux de meilleure qualité (stables, forts et plus discriminants) que d'autres, nous avons considéré que les caractéristiques neurophysiologiques des utilisateurs pouvaient également être des prédicteurs pertinents de la performance des BCI. De tels prédicteurs pourraient être accessibles en utilisant le signal EEG au repos avant le début de l'entraînement. Dans les études précédentes, certains prédicteurs ont été identifiés. Cependant, ils ont été révélés dans des expériences uniques, chacune incluant un petit nombre de sujets et de sessions. De plus, ils correspondent à des modèles univariés et les études ne montrent pas dans quelle mesure un prédicteur donné pouvait réellement prédire les performances BCI de sujets non vus. Or, pour être utiles, ces modèles devraient être stables et précis, c'est-à-dire être capables de prédire réellement les performances BCI de nouveaux utilisateurs. Ces modèles doivent également prendre en compte plusieurs variables et être généralisables à travers les expériences et les ensembles de données.

L'objectif principal de cette partie est d'étudier les prédicteurs potentiels des performances d'utilisateurs de TM-BCI, en vue d'améliorer leur entraînement sur la base de nos résultats. Notre but est de créer des modèles computationnels de d'entraînement des utilisateurs de BCI qui pourraient prédire les performances de divers utilisateurs de BCI, en fonction de leurs traits, de leurs compétences et de leurs caractéristiques neurophysiologiques. En effet, comme ces caractéristiques sont accessibles avant le début de l'entraînement BCI, être capable de trouver un modèle qui prédit les performances des TM-BCI en fonction de ces caractéristiques pourrait nous permettre de sélectionner les (potentiels) utilisateurs les plus réactifs et/ou de concevoir des protocoles d'entraînement adaptés au profil des

utilisateurs et optimisant ainsi les capacités d'apprentissage.

Les caracteristiques issues de la machine

Les BCI identifient l'intention des utilisateurs en analysant leur activité cérébrale. Les utilisateurs peuvent alors interagir avec le monde extérieur sans aucun mouvement [16]. Cette activité est le plus souvent mesurée à l'aide d'un EEG, puis traitée à l'aide de différentes méthodes à différents niveaux : le niveau de prétraitement, le niveau d'extraction des caractéristiques discriminantes et enfin le niveau de classification. Par conséquent, idéalement, lorsqu'il apprend à contrôler un TM-BCI, l'utilisateur apprend également à générer des signaux cérébraux stables et distincts pour chaque classe (c'est-à-dire chaque tâche) [**?**]. Pour pouvoir le faire, les caractéristiques extraites des signaux EEG enregistrés doivent être cohérentes. En d'autres termes, elles doivent correspondre aux modèles neurophysiologiques identifiés dans la littérature. Au cours de la phase de calibration d'un système, des méthodes d'apprentissage discriminantes basées sur les données sont généralement utilisées pour effectuer l'extraction des caractéristiques EEG. Ces caractéristiques sont ensuite utilisées pour distinguer des classes, à l'aide d'algorithmes de classification. Une technique de traitement du signal populaire pour les TM-BCI basés sur l'EEG est l'algorithme Common Spatial Pattern (CSP), qui apprend les filtres spatiaux qui discriminent le mieux deux classes de TM [17, 18]. Les filtres CSP maximisent la variance du signal filtré spatialement dans une condition et la minimisent pour l'autre classe. Avant d'utiliser les filtres CSP, plusieurs paramètres doivent être sélectionnés : le filtre passe-bande, l'intervalle de temps et le nombre de filtres à utiliser. Lors de la calibration du système, il est courant d'utiliser des paramètres génériques (c'est-à-dire les mêmes paramètres) pour tous les sujets, par exemple la bande passante standard de 8-30 Hz pour les TM-BCI [17]. Pourtant, le choix individuel de ces hyperparamètres pour la classification basée sur CSP pourrait améliorer les performances en ligne [19]. Certaines méthodes basées sur les données ont été développées pour sélectionner des hyperparamètres spécifiques à chaque sujet [18].

Bien que ces algorithmes d'extraction de caractéristiques et de classification soient couramment utilisés et aient prouvé leur efficacité, ils sont presque exclusivement axés sur les données. En effet, ils n'incluent que très peu d'antécédents neurophysiologiques et, font plutôt confiance aux données EEG (potentiellement bruitées) enregistrées lors de la phase de calibration. Dans cette partie, nous émettons l'hypothèse que, pour être plus efficaces, ces méthodes devraient prendre en compte certaines contraintes qu'il est nécessaire d'identifier. En effet, toutes les propriétés des caractéristiques de l'utilisateur ou de la machine appris à partir des données BCI ont-elles la même probabilité d'être associées à de bonnes performances en pratique ? Si ce n'est pas le cas, quelles propriétés sont plus souvent associées à des performances supérieures ? Est-ce que prendre en compte ces propriétés dans les algorithmes d'apprentissage automatique des BCI conduirait à de meilleures performances de décodage en pratique ?

Pour répondre à ces questions, nous étudions dans un premier temps l'impact des propriétés de la bande de fréquence la plus discriminante (MDFB) [18] sélectionnée par les algorithmes d'apprentissage automatique sur les performances en ligne. L'algorithme MDFB vise à sélectionner la bande de fréquence qui discrimine le mieux deux classes. Dans un second temps nous nous concentrons sur les caractéristiques extraites de l'algorithme CSP et sur un algorithme de classification fréquemment utilisé dans les BCI : l'analyse discriminante linéaire (ADL).

Tout d'abord, l'objectif était d'étudier la relation entre les performances des BCI et les caractéristiques de la bande de fréquence la plus discriminante (MDFB) spécifique à chaque sujet, sélectionnée par un algorithme heuristique populaire [18]. Pour cette étude, nous avons utilisé les données de 80 participants provenant de deux expériences différentes (mais similaires en termes de protocole) dans lesquelles les participants devaient apprendre à contrôler un TM-BCI à deux classes. Nous avons d'abord cherché à savoir s'il existait une relation entre les caractéristiques de la MDFB choisie et les performances en ligne des sujets (c'est-à-dire la précision de la classification). Ensuite, nous avons étudié la causalité de leur relation, et enfin si nous pouvions trouver des paramètres qui pourraient être utilisés pour affiner l'algorithme pour l'optimiser. Nos analyses ont suggéré qu'il

existait une corrélation entre les caractéristiques de la bande de fréquences sélectionnée spécifique à l'utilisateur, c'est-à-dire la valeur moyenne et la largeur de la MDFB sélectionnée, et la précision de la classification. Pour étudier un éventuel lien de causalité entre les deux, nous avons ajouté des contraintes à l'algorithme, en imposant des caractéristiques associées à des performances plus élevées (sur la base de nos premiers résultats), et nous l'avons utilisé dans une expérience spécialement conçue pour cette étude (Dataset B). Nous avons ensuite comparé les performances en ligne obtenues pour les sujets de ce groupe expérimental (utilisant l'algorithme contraint) avec les performances en ligne de sujets appariés dans un groupe de contrôle (utilisant l'algorithme non contraint, c'est-à-dire l'algorithme original). Cette étude ne nous a pas permis de déterminer un lien de causalité. En effet, si les performances moyennes en ligne obtenues avec l'algorithme contraint étaient légèrement supérieures à celles obtenues avec l'algorithme non contraint, cette différence n'était pas significative. Une des raisons pourrait être que l'ajout de contraintes pourrait être bénéfique pour certains sujets mais néfaste pour d'autres. Par conséquent, dans une dernière partie, en plus de construire des modèles pour prédire la performance des utilisateurs de TM-BCI en utilisant des paramètres extraits des algorithmes d'optimisation utilisant la MDFB, nous avons essayé de construire un modèle de classification qui pourrait choisir l'algorithme optimal (contraint ou non contraint) pour chaque sujet. Nous avons été en mesure de prédire les performances des utilisateurs en utilisant des paramètres extraits des algorithmes d'optimisation MDFB. Pour construire nos modèles, nous avons ajouté trois paramètres en plus de la *largeur* de la MDFB et de la *moyenne* et avons donc proposé trois prédicteurs potentiels des performances des TM-BCI. Pour identifier ces prédicteurs, nous avons modélisé chaque signal de corrélation comme une fonction gaussienne de la fréquence et utilisé les paramètres du modèle gaussien comme prédicteurs (c'est-à-dire sa moyenne $f0$ et sa variance b et la valeur maximale de la courbe$a$). Les modèles de prédiction comprenaient trois facteurs : i) la *largeur de la MDFB* et ii) la valeur maximale du modèle gaussien a avec des poids forts par rapport aux autres facteurs, et iii) la valeur moyenne du modèle gaussien $f0$, avec un poids négatif plus faible. Ce résultat a renforcé nos résultats précédents, c'est-à-dire que les sujets ayant une *largeur de MDFB* plus élevée ont tendance à avoir de meilleures performances que les sujets ayant une *moyenne de MDFB* plus faible et les sujets ayant une *moyenne de MDFB* supérieure à 16 Hz (dans la bande β) semblent avoir de moins bonnes performances. En outre, nous avons pu trouver un nouveau prédicteur des performances du MI-BCI : la valeur maximale de la fonction gaussienne optimisée et sa moyenne. Dans une dernière partie, nous avons décidé d'utiliser ces coefficients pour choisir un algorithme spécifique au sujet (contraint ou non contraint) afin d'augmenter éventuellement les performances en ligne. Nos résultats ont révélé que l'utilisation d'un algorithme contraint pouvait aider à améliorer les performances en ligne des sujets ayant soit des signaux EEG distincts, soit aucun signal EEG distinct, et une fréquence f0 dans la bande basse β. Ce résultat souligne l'intérêt d'utiliser des prédicteurs des performances des BCI afin de choisir l'algorithme d'apprentissage automatique optimal pour chaque sujet. Par exemple, pour les sujets présentant des modulations SMR très faibles et une fréquence discriminante dans la bande basse β, l'utilisation de valeurs par défaut (ici nos contraintes), pourrait être meilleure que de laisser l'algorithme apprendre à partir des données, car il n'y a rien de clair dans les données pour apprendre. Cependant, pour les sujets présentant de fortes modulations SMR, il pourrait être plus intéressant de laisser l'algorithme apprendre à partir des données.

Dans un second temps, nous avons étudié d'autres algorithmes d'apprentissage automatique utilisés dans les BCI (c'est-à-dire les algorithmes CSP et ADL) afin d'identifier les caractéristiques des filtres spaciaux et des classifieurs pouvant conduire à de meilleures performances. Pour cette étude, nous avons utilisé les données de 82 participants provenant de deux expériences différentes dans lesquelles les participants devaient apprendre à contrôler un TM-BCI à deux classes. Nos analyses ont suggéré deux modèles basés sur un ensemble de données, significativement meilleurs que le hasard et suffisamment fiables pour prédire les performances de nouveaux sujets provenant d'une autre expérience (qui avait le même protocole). Le premier modèle, c'est-à-dire le modèle 1, comprenait des caractéristiques extraites des filtres et des motifs des CSP et des motifs du ADL. Le deuxième modèle, c'est-à-dire le modèle 2, était basé sur les caractéristiques spécifiques des sujets extraites des motifs et des filtres CSP pondérés par les poids normalisés les motifs du ADL. Les facteurs ayant le

plus d'impact dans notre modèle étaient liés à l'importance des poids des motifs du ADL (c'est-à-dire la moyenne des poids des motifs du ADL dans le modèle 1), à la latéralité des filtres CSP (latéralité des filtres CSP, distance des filtres CSP D1 dans le modèle 2), à la latéralité des motifs CSP (distance $\|v_{max} - v_{c3orc4}\|^2$ dans les modèles 1 et 2) mais aussi à la distorsion des motifs CSP (rapport R dans les modèles 1 et 2). Comme on s'attend à une latéralisation des signaux cérébraux produits par l'utilisateur lorsqu'il effectue une TM de la main gauche et de la main droite, les facteurs inclus dans nos modèles peuvent donner des informations sur la capacité de l'utilisateur à produire de tels signaux latéralisés. De plus, nous avons observé que le fait d'avoir un facteur de l'ALD avec un poids moyen proche de zéro a un impact important lors de la classification car les sujets peuvent avoir des difficultés à trouver une stratégie fiable et donc à contrôler une BCI. Une prochaine étape serait d'ajouter des contraintes à ces algorithmes en fonction de nos résultats (par exemple, imposer une latéralité dans les filtres CSP) ou de trouver une métrique permettant de visualiser plus facilement la principale source d'information et d'adapter notre formation en conséquence (par exemple, une formation en neurofeedback pourrait aider certains sujets à obtenir une meilleure latéralisation des signaux cérébraux). Dans l'ensemble, nos travaux ont révélé que l'étude des caractéristiques extraites des méthodes axées sur les données pourrait être intéressante pour mieux comprendre pourquoi certains sujets ont des difficultés à contrôler un BCI. Cela pourrait également donner une indication sur la voie à suivre à l'avenir pour adapter et améliorer ces méthodes en ajoutant des contraintes supplémentaires pertinentes.

Contents

1

Introduction

The pre-history of brain–computer interfaces (BCIs) started with the first recording of human electroencephalography (EEG) in 1924 by a German physiologist and psychiatrist named Hans Berger [20]. H. Berger followed the pioneering work done by Richard Caton (1842–1926) in England with animals [21]. Using the EEG, he was the first to describe the different waves, or rhythms, that were present in the *normal* and *abnormal* brain, and to use the terms alpha and beta waves to describe specific activities. His method consisted in inserting silver wires under the patients' scalp, one at the front of the head and one at the back. Later, he used silver foil electrodes attached to the head by a rubber bandage.

Neurofeedback methods were first developped in the late fifties and used to train people to voluntarily control specific physiological states. For intance, Razran proposed the concept of biofeedback [22] suggesting that through the use of physiological instruments, people can observe their physiological states and learn to control them. Joseph Kamiya oriented his researches on human brain wave frequencies and suggested that brain waves are not involuntary states, but can be controlled by humans [23]. In his first experiment, he asked the participants to tell, at different moments, if they were in "alpha state" or not. In other words, he asked them to determine whether their alpha amplitude was high or low. At the beginning, all were at chance level, but after a while, some of them were able to answer correctly with performances better than chance. In a second experiment, Kamiya asked participants to enter or quit alpha state upon the elicitation of specific clues (one or two bips). After a certain training, some participants succeded in voluntarily increasing or decreasing their alpha amplitude It was the beginning of neurofeedback trainings (NFT).

In 1965, although the term BCI had not yet been coined, the American composer Alvin Lucier used EEG and analog signal processing hardware to stimulate acoustic percussion instruments and perform the piece Music for Solo Performer also referred to as the 'brain wave piece' [24]. To perform the piece one must produce alpha waves and thereby 'play' the various percussion instruments via loudspeakers that are placed near or directly on the instruments themselves. This example well describes scientists' fascination for the direct use of human brain activity to control objects or to communicate.

With the development of computer sciences in the 1970s, the scientific community started to imagine systems enabling humans to communicate using their brain activity alone. At first, it was simple applications using, for example, the users' alpha waves amplitudes. In a paper published in 1973, UCLA Professor Jacques Vidal coined the term "Brain-Computer Interfaces" and produced the first peer-reviewed publications on this topic [25]. BCIs were later defined in 2002 by Wolpaw et. al, in a reference paper entitled "Brain-Computer Interfaces for Communication and Control" [16] as a communication system in which messages or commands that an individual sends to the external world are created without using peripheral nerves and muscles, i.e., by using brain signals alone. In the same paper, the importance of developing and learning a new skill to control a BCI was also addressed. Such a skill involves voluntary control of specific electrophysiological signals.

Since the first brain-computer communication systems, advances have been made in the field of computer sciences, neurosciences and machine learning and, in the same vein, BCIs have undergone exponential development and great diversification. Currently, BCIs can be classified into three main categories: Active, Reactive and Passive [26].

- **Active BCIs** use brain activity that is directly consciously controlled by the user, independently from external events, for controlling an application. Active BCIs can be divided into two sub-categories. The first one is BCIs based on slow cortical potentials (SCPs). They rely on the voluntary production of negative and positive potential shifts and was first introduced in 1990 by Birbaumer et. al. [27]. SCPs are defined as Event-Related Potentials (ERPs) that are triggered, from the upper layer of the cortex, from 3 seconds to several seconds after the eliciting

event. ERPs are automatically generated by the brain in response to a stimulus. Such stimulus can be of exogenous origins (reaction to a presented external stimulus) or endogenous origins (reaction to expectation of stimuli). In the case of SCPs, it is a stimuli from endogenous origin. Apart from SCPs, there are many other EEG-based BCI systems that rely on sensorimotor rhythms (SMRs), which are brain rhythms related to motor actions over sensorimotor areas [28] and allow anatomically specific voluntary regulation. The user is asked to perform specific motor-imagery (MI) tasks (such as motor-imagery of the limbs). When they are being performed, these mental tasks induce an Event-Related Desynchronisation (ERD) [29, 30]. Once the user stops performing the task, an Event-Related Synchronisation (ERS) in induced. Once the system is able to recognize the brain pattern associated with a specific task, this pattern is linked to a specific command. In this type of BCI, for instance, imagining movements of a hand can be considered as a command for external devices [30]. Using the same approach (i.e., ERD/ERS induced in specific brain areas), it is also possible to use different mental tasks (MT) such as mental calculation or mental rotation. We then talk about MT-BCIs. Finally, it has been demonstrated that MT-BCI systems can be used to enable people with motor disabilities to operate an environmental control system or a simple word-processor faster than SCP-BCI systems [31, 32]. It has been also proposed as a promising technology for both impaired and healthy users. For impaired people BCIs can be used as assisting technologies, for example, wheelchairs (see Figure 1.1) or neuroprosthetics [33] and for healthy users, for instance, BCIs can be used to control video games or virtual reality applications [34].

Figure 1.1: Controlling a wheelchair using motor imagery (picture from EPFL)

– **Reactive BCIs** use brain activity arising in reaction to external stimulation, which is modulated by the users' attention, for controlling an application. This particular cerebral activity are ERPs. A commonly used paradigm for reactive BCIs is the P300 setting [35], first introduced by Farwell and Donchin in 1988 [36], which extracts outputs from brain activities in reaction to stimuli on which users focus their attention (see Figure 1.2). The stimuli may be characters or icons displayed on a screen. They can also be spatially oriented sounds or vibrations on different body parts. Each stimulus is highlighted at specific, pre-determined moments and the user is asked to focus their attention on the stimuli he wants to select. The P300 is a positive cortical potential that appears around 300ms as a result of an individual voluntarily focusing attention on a specific external stimulus. Reactive BCIs do not require a long training but require a high level of attention making it difficult to use in interactive contexts with multitasking [37].

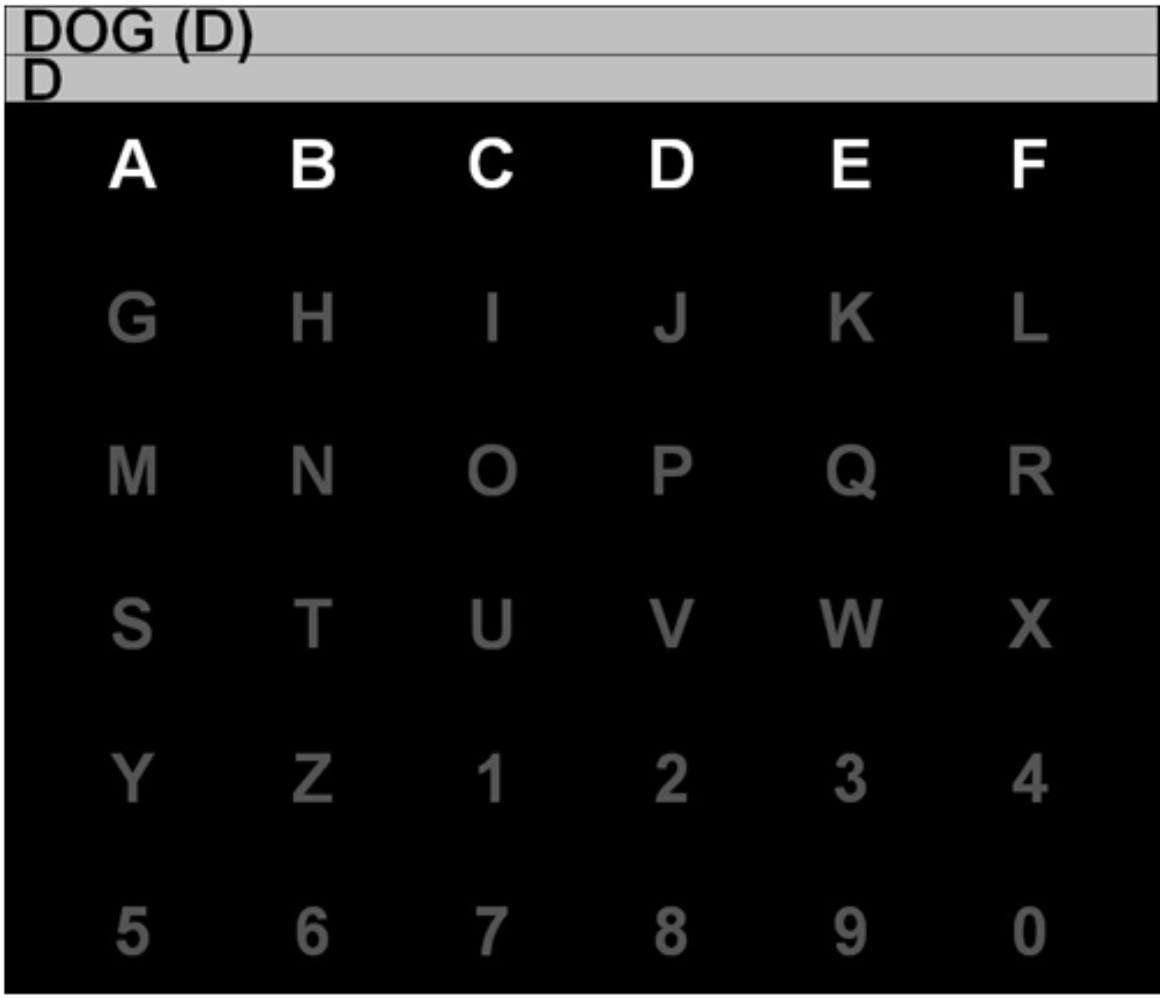


Figure 1.2: The P300 speller interface, as displayed on the user screen. To spell the word "DOG", as rows and columns flash successively. When a column or row contains the character a subject desires to communicate, the P300 response is elicited.

– **Passive BCIs** use spontaneous brain activity without the purpose of voluntary control [26]. The user's mental state is measured in order to adapt an application/interface accordingly. For example, a passive BCI has been developed for the assessment of mental workload of air traffic controllers [38] (see Figure 1.3). As there is no voluntary interaction between the user and the machine, they may not be considered as BCIs. However, if we consider the definition of BCI of Wolpaw [39], i.e., "a system that measures central nervous system (CNS) activity and converts it into artificial output that replaces, restores, enhances, supplements, or improves natural central nervous system output and thereby changes the ongoing interactions between the CNS and its external or internal environment", BCIs are systems that measure brain activity and convert it into artificial output that reflect the user's intent or other aspects of current brain function (e.g., state of alertness). Hence, passive BCI can be considered as BCIs.

As the object of this thesis is to contribute to the improvement of BCIs dedicated to communication and control in order to render them more usable and accessible for patients as well as for the general public, we focus on active BCIs and more specifically on MT-BCIs using EEG.

The design of an MT-BCI system is a complex task that requires multidisciplinary skills such as computer science, signal processing, neurosciences but also psychology. When a user is asked to control an MT-BCI, two phases are generally required: an offline training phase to calibrate the system (i.e., training of the system) and an online phase during which users learn to generate stable and distinct signals to control the BCI (i.e., training of the user). During the offline calibration

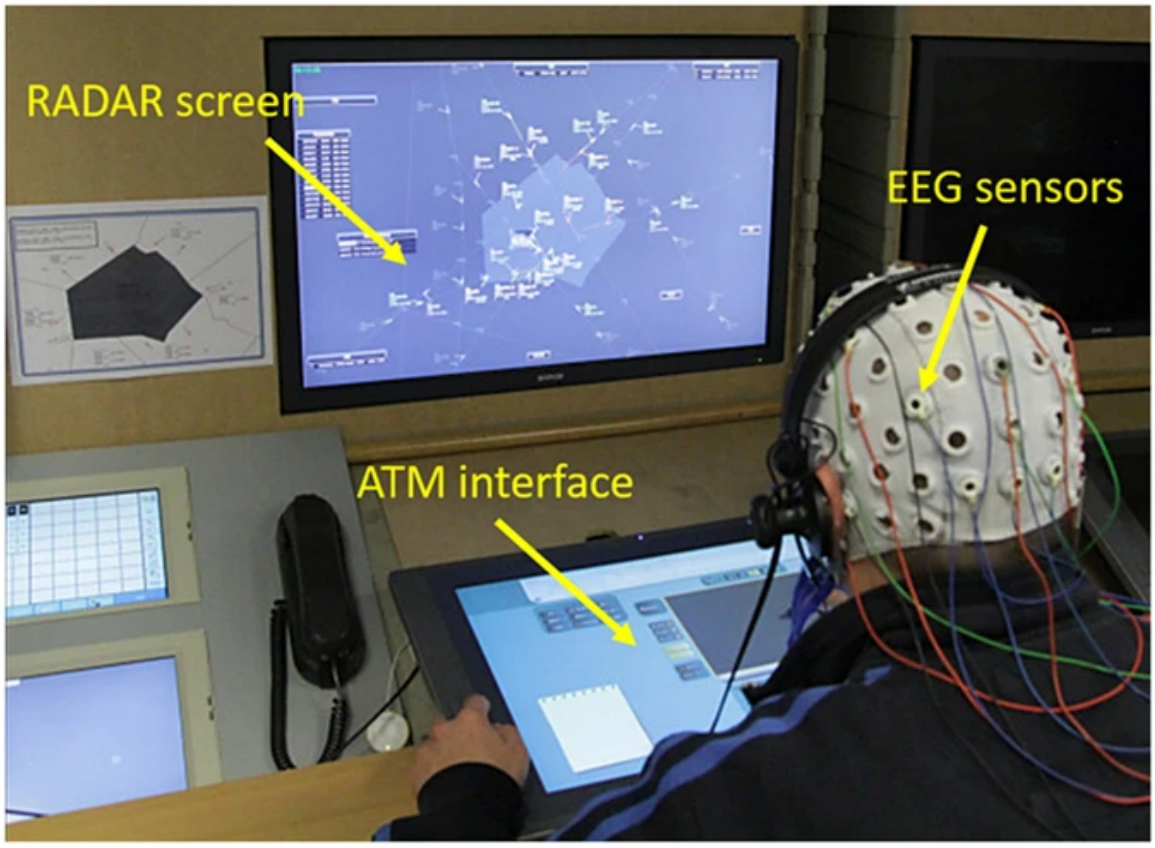


Figure 1.3: Experimental setup as in [38]: The air traffic controller's brain activity has been recorded continuously during the execution of the air traffic management scenario

phase, spatial, temporal and spectral features are used to classify different brain patterns and thus discriminate MT-related patterns. Different discriminative learning methods exist for EEG feature extraction in MT-BCIs. For instance, spatial filters such as common spatial pattern (CSP) [40] or independant component analysis (ICA) [41] generally unmix the signals to recover the original sources. Then, the final phase is to translate the features into commands. To do so, classification algorithms [42, 43] are used to automatically assign a class (corresponding to an MT) to the feature vector previously extracted. Therefore, during the online training phase the user is trained to generate brain signals that the system, previously calibrated, can associate with a specific MT. Therefore, users can adapt their strategy [12] (for example timing, perspective, along with dimensional, emotional or sensory characteristics of the task) so that the MT they are performing gets recognised as often and as well as possible by the system. As, machine learning is coupled with a user training phase to achieve acceptable control, MT-BCIs are called co-adaptive communication systems.

The online MT-BCI system requires to follow a closed-loop interactive process (see Figure 1.4) for which commands are sent through the modulation of specific brain patterns. The process is composed of five steps: instructions, execution of a MT, signal recording, signal processing and feedback [44].

1– **Instructions:** this step consists in giving information about how the interaction should be done or the system operated [12].

2– **Mental task:** In order to modulate their brain patterns (and then send the associated commands), users perform mental tasks. For instance, motor imagery (MI) of the left hand is underlain by modulations of the SMRs over the right hemisphere [28]. Therefore, in order to send different commands, users can perform different MT.

3– **EEG signal recording:** this step consists in using EEG sensors in order to obtain signals that reflect the user's brain activity [45].

4– **EEG signal processing:** this step encompasses three stages: preprocessing (cleaning and denoising the input data), feature extraction (describing the signals by a few relevant features) [46] and classification (assigning a class to a set of features) [47].

5– **Feedback:** this step provides the user with a feedback about the identified MT. This feedback helps the user to learn to control the system.

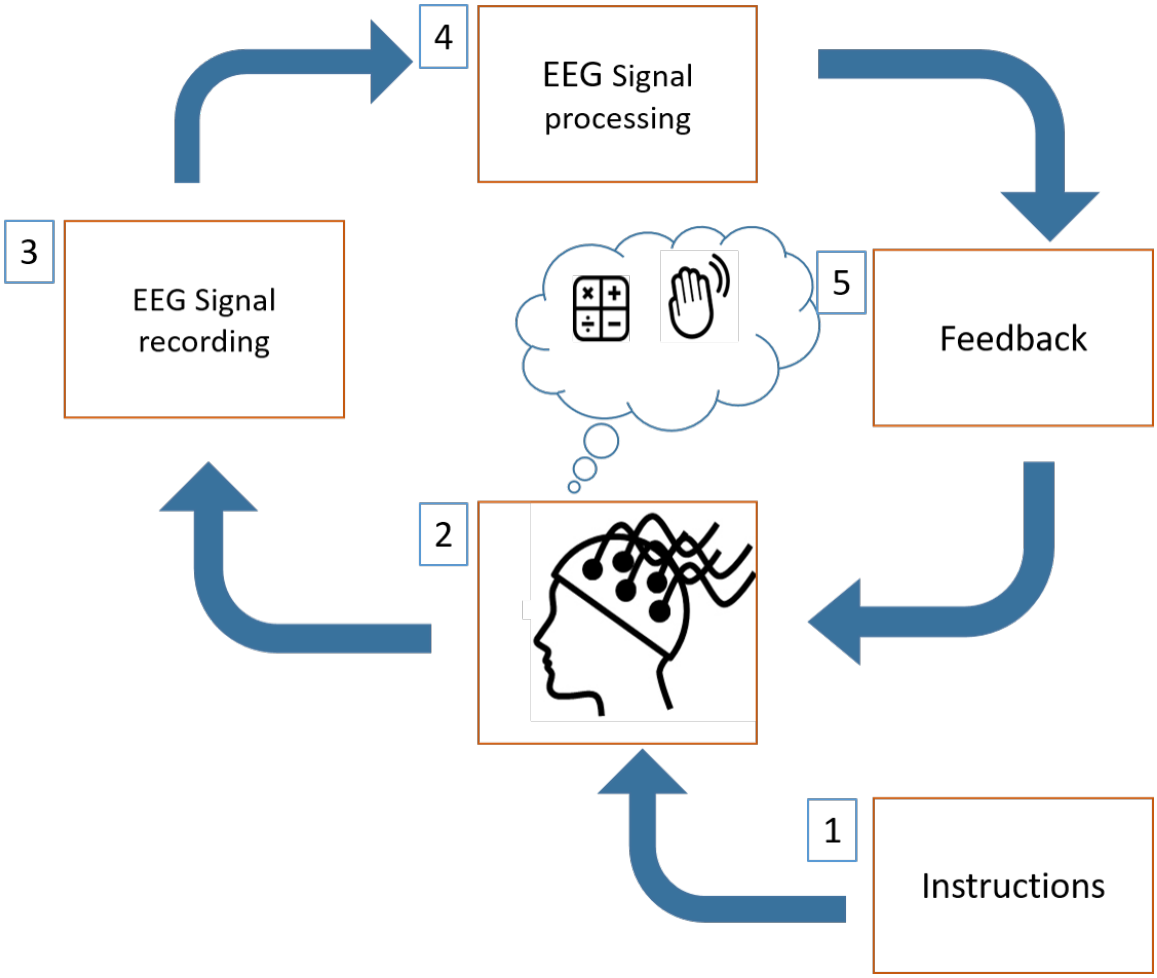


Figure 1.4: Online BCI loop

It is important to note that EEG signals are subject-specific. The high between-subject variability currently requires BCI systems to be calibrated and adapted to each user. The training dataset should contain EEG signals recorded while the subject performed each MT several times, according to given instruction in order to smooth the within-subject variability, which can also be high, both within and between training sessions.

Among the different training approaches for MT-BCIs [11, 48, 49], one of the most widely used is the standard Graz training protocol [11, 49]. A large majority of other existing MT-BCI training protocols are variants of the Graz training protocol, as they use similar timings, feedback and training tasks [12, 50–52]. This approach is organised in two stages. First training the system and second training the user. During the first stage, the user is instructed to successively perform a certain series of MTs. From the recorded EEG signals collected during the different MTs, the system extracts characteristic EEG patterns that are specific to each MT. These extracted patterns are then used to train a classifier, the goal of which is to determine the class to which the EEG signals belong. For MT-BCI training protocols that last over several sessions, performed on different days, it is common to regularly retrain the classifier on newly acquired data in order to take into account cap variations and the EEG non-stationarity due, in part, to the condition/state in which the user is (which can

change from one session to another). The second stage consists in training the user. To do so, the user is instructed to perform the same MT tasks, but this time feedback (provided by the classifier, which was optimised during the first stage) is provided to inform the user which MT the system has recognised and how confident the system is that the task it has recognised is the one being performed by the user. Thus, the goal of the user will be to find strategies and adapt the way they perform the MTs such that the system recognises them as well as possible. Chapter 4 details the standard Graz protocol.

MT-BCIs are promising technologies for numerous applications, e.g., for assistive technologies (e.g., motor substitution and recoverty) [33, 53, 54] or entertainment (e.g., video games) [34], but are unfortunately not yet reliable enough and remain barely used outside laboratories [55]. One of the reasons is the limitations due to the brain recording techniques used, here the EEG. Indeed, most of the devices that are used in laboratories are too cumbersome (due to the amplifiers and the wires) to be considered as portable and use wet-type EEG sensors that need conductive gel and sometimes skin abrasion which make them uncomfortable for the user. Furthermore, those types of sensors are inappropriate for long-time measurement (more than 1 or 2 hours) and the quality of the signal thus decreases over time because of the drying of conductive gel. In addition, the low signal-to-noise ratio due to the sensitivity of the sensors and to the EEG method itself (large diffusion) make the spatial resolution rather low, i.e., 22 to 37 cm^3 for a conventional 19-electrode systems [56]. Still, most systems are relatively inexpensive and can have a high temporal resolution and dry electrodes are increasingly used. Another issue is the software limitation that can cause an erroneous recognition of the executed mental commands. One of the current challenges is to design more reliable signal processing algorithms to increase the reliability of BCIs. Yet, even when most of the research efforts are focusing on those two last points, BCI performances still need to be improved and BCIs are still unreliable [5]. The main measure of this BCI unreliability is the large proportion of users (15% to 30%) who do not seem able to produce brain activity patterns that the computer can discriminate while performing MTs. Such users are commonly called "BCI illiterate" [4] [i]. This issue is less studied but not less important. Indeed, if the user cannot generate "understandable" signals (i.e. stable and distinct brain signals for each task), they will not be able to control the system even with the best electrodes and algorithms. Producing such brain signals is a skill that must be acquired and mastered by the user [58]. Recent research results have actually shown that the way BCI users are currently trained was suboptimal, both theoretically [6, 7] and practically [8]. One reason could be that each user starts with different skills, brain activity and traits of personality but with the same training protocol. Moreover, the user is known to be one of the main cause of EEG signals variability in BCI, due to his/her change in mood, fatigue, attention, etc. [9]. Therefore, adapting the training procedure to each user is a promising area of research for improving the reliability of BCIs. It is in this context that we consider as essential to understand MT-BCI learning and identify factors influencing it.

[i] There is valid criticism about the broad use of the term "BCI illiteracy". Indeed, this concept seems to rely on an inaccurate assumption that users and their traits are the sole responsible for poor BCI performances [57].

2

Objectives and contributions

In the following chapter, we first explain the rational of our research challenge. We argue that to make BCIs more reliable, efficient and accessible, it is necessary to redefine training protocols and more precisely to adapt them to each users. In this context, we will define the objectives of this PhD thesis and introduced the different scientific contributions.

Contents

2.1 Thesis Objectives

Since current training protocols are suboptimal, it appears necessary to redefine them so that they comply with theoretical guidelines such as the ones introduced in Lotte et. al. [6] concerning training environment, instructions, tasks and feedback. For instance, standard feedback cannot quantify specifically and uniquely the BCI users' skills, i.e., how well the user can self-modulate their brain activity to control the BCI [8].

In order to improve and adapt users' training, it is necessary to understand the factors impacting the MT-BCI user training process and to improve the training protocols accordingly. Commonly called predictors, those factors that can explain and predict MT-BCI performances can be gathered in four main categories, as depicted in Table 2.1, all related to the user:

- Factors related to the users' traits, defined as stable and long lasting users' characteristics with respect to one's environment [59]. For instance personality profile, cognitive profile and demographic characteristics are part of this category.
- Factors related to the user experience. For instance experience with musical instruments or video games.
- Factors associated with the states of the user. Users' state are described as "temporary, brief and caused by external caused" [59]. Such factors could be attention, motivation or mood.
- Factors related to the users' neurophysiological characteristics or in other words, the users' ability to produce brain signals of interest.

More details about factors related to users' trait and factors related to the users' neurophysiological characteristics will be given in Chapter 5 and Chapter 6 respectively where a review of the literature will be proposed.

In addition, while often presented as promising assistive technologies for motor-impaired users, MT-BCI remain barely used outside laboratories due to low reliability in real-life conditions but also high between-session variability. Finally, as BCIs are communication systems between a *user* and a *machine*, we believe that predictors of BCI performances should not only be related to the users' characteristics but also characteristics of the machine (properties of the classifier and signal processing approaches).

Therefore, among the numerous possible improvements that can be made in the research of MT-BCI predictors to adapt users' training, we are going to address three main points in this PhD thesis: 1) designing and studying a multi-class MT-BCI for longitudinal training (20 sessions over 3 months) of an end-user (i.e., tetraplegic user), 2) creating computational models of BCI user training that could predict the performances of various BCI users, over training time, based on their traits, skills, neurophysiological characteristics but also on characteristics of the machine , 3) studying the generalization of the predictive models across experiments and datasets.

2.1.1 Long-term BCI training of a tetraplegic user

Current BCI systems are mostly studied and evaluated inside laboratories, with a restricted number of sessions, in highly controlled conditions and with neurotypical users'. To further develop BCI technology, it is necessary to study BCI in real-life or close to real-life conditions, with end-users, e.g., severely motor-impaired ones. It is also essential to understand the between-session variability but also the user learning in those conditions. Indeed, beside the largely unknown phenomena in the activity of neuronal populations which lead to non-stationarity of EEG signal ([77]), some variability sources including various environmental noises and changes in users' mental states such as their attention, fatigue or stress level are expected in real-life applications. As in this thesis manuscript we focus on studying and improve BCI training protocol, the study of a long-term BCI training of a tetraplegic user could be used to better understand BCI users' learning and between-session variability.

Table 2.1: Summary table of different factors which have been related to MT-BCI performance in the literature.

TRAITS	PERSONALITY	• Locus of control for dealing with technology [60] • Apprehension, self-reliance and Visual/Verbal learning style [52]
	COGNITIVE PROFILE	• Attention span [61] • Attentional abilities [62] • Memory span [62] • Visual-motor coordination [61, 63] • Spatial abilities [52, 64, 65] • Kinesthetic and visual-motor imagination [66]
	DEMOGRAPHIC DATA	• Age [67] • Gender [67] • Hendedness [68]
STATES	EMOTIONAL STATE	• Mood []
	COGNITIVE STATE	• Attention level [69] • Motivation [61, 70, 71] • Mastery confidence [71] • Fear of BCIs [60, 72] • Control beliefs [72] • Fear of incompetence [71] • Sefl-efficacy [70]
EXPERIENCES	PERSONAL EXPERIENCE	• Playing music instrument [67] • Practicing sports [67] • Paying video-games [67] • Time spent typing [73] • Hand, arm and full body movements [73] • Consuption of affective drugs [73]
NEUROPHYSIOLOGICAL CHARACTERISTICS	COGNITIVE STATE	• Attentional level [74] • Attentional processes [75] • Memory performance [76]
	INDICATOR OF SIGNAL STRENGH	• Sensory motor rythms strengh [**?**]

2.1.2 Computational models of BCI user training

To our knowledge, most of the identified factors associated to BCI performances (see Table 2.1) have been revealed in single experiments, each including a small number of subjects and sessions; or they correspond to univariate models (i.e., including a single factor). Moreover, such univariate models were often correlations only, and did not study how well a given factor (or predictor) could actually predict the BCI performances of unseen subjects. These limitations might explain, at least in part, why most of the highlighted factors have not been replicated since. Designing interpretable computational models of BCI user training based on multiple variables would allow the identification of factors impacting MT-BCI performances and provide optimal instructions, feedback, training exercises or calibration methods for each users. In other words, we need to understand which users' traits, skills, neurophysiological patterns and which characteristics of the machine (i.e., properties of the classifier and signal processing approaches) impact MT-BCI performance and how these factors interact to influence them through exercises or experimental design.

2.1.3 Generalization of the predictive models across experiments and datasets

Currently, a large majority of identified predictors do not generalize across experiments and datasets. Indeed, too often, predictors of MT-BCI performances are found on experiments using either a low number of subjects or only one MT-BCI session. When using a low number of subjects or sessions, within-subject variability could play a large role in the obtained performances as subjects performances vary over the course of individual session and their skill in operating a BCI does not remain constant over the course of a session [78]. Therefore, it seems necessary that reliable predictors should generalize across experiments with a similar but also a different protocols.

2.2 Approach and contributions

This manuscript describes the work we carried out in order to address the three objectives mentioned above. More precisely, the first chapter of this thesis (Chapter 3), and our first scientific contribution, is an experimental study of learning through the participation of our research team in the CYBATHLON competition in 2019. We designed and studied a multi-class MT-BCI for longitudinal training (20 sessions over 3 months) of a tetraplegic user for the CYBATHLON BCI series 2019. Then, the first part of the manuscript is dedicated to the Materials and Methods used in the different additional scientific contributions we proposed. These contributions can be gathered into two parts: Part I, focusing on the identification of user-related factors used to predict MT-BCI performances, Part II, in which we investigated the machine characteristics (i.e., characteristics of what the machine has learned). More specifically, in Part II we studied characteristics from feature extraction and classification algorithms. More details are given in the following sections.

2.2.1 Contribution 1: Long-term BCI training of a Tetraplegic User

While promising, MT-BCIs are generally not yet reliable enough for being used in practice, i.e., in *real-world* applications, outside laboratories. Indeed, it appears that the decoding of users' mental commands is subject to high error rates [**?**, **?**]. There is thus a need to design long-term reliable BCIs that can be used outside-of-the-lab by end-users, e.g., severely motor-impaired ones. Therefore, in Chapter 3 we propose and evaluate the design of a multi-class Mental Task (MT)-based BCI for longitudinal training (20 sessions over 3 months) of a tetraplegic user for the CYBATHLON BCI series 2019. In this BCI championship, tetraplegic pilots are mentally driving a virtual car in a racing video game. We aimed at combining a progressive user MT-BCI training with a newly designed machine learning pipeline based on adaptive Riemannian classifiers shown to be promising for real-life applications. We followed a two step training process: the first 11 sessions served to train the user to control a 2-class MT-BCI by performing either two cognitive tasks (REST and

MENTAL SUBTRACTION) or two motor-imagery tasks (LEFT-HAND and RIGHT-HAND). The second training step (9 remaining sessions) applied an adaptive, session-independent Riemannian classifier that combined all 4 MT classes used before. This classifier aims to address the between-session variability. Indeed, Riemannian classifiers generally reach high accuracy, making real-life BCI applications promising. Moreover, as our Riemannian classifier was incrementally updated in an unsupervised way it would capture both within and between-session non-stationarity. Importantly, we also report on the evolution of the user's neurophysiological patterns and user experience throughout the BCI training and competition to better understand the learning of the user.

2.2.2 Contribution 2: Models based on user-related factors

BCI control is a skill that must be acquired and mastered by the user [10–12]. Indeed the efficiency of MT-BCIs inherently depends on the users' ability to successfully encode mental commands into their brain signals. In other words, it relies on the users' ability to produce EEG patterns that are stable each time they intend to send one same command, and distinct between the different mental commands. These characteristics of stability and distinctiveness are essential for mental commands to be efficiently decoded by the system [79]. Indeed, if the user cannot produce stable and distinct EEG patterns, then no machine learning algorithm would be able to extract reliable commands. Hence, it seems essential to understand the cognitive and neurophysiological processes that underlie the ability to efficiently encode mental commands through the performance of mental imagery tasks, i.e., to produce stable and distinct EEG patterns. Moreover, it is estimated that 10 to 30% of potential BCI users [14, 15, 80, 81] would not be able to control current MT-BCI applications at all, at least with current BCI systems and for naive users. To make MT-BCIs more reliable, researchers have mainly focused on hardware (e.g., electrodes) and software (e.g., signal processing algorithms) improvements, but less on the improvement of user-training procedures. Yet, as previously explained, this aspect is also essential. Indeed, individuals vary in their personal characteristics across many dimensions. Such characteristics include person's demographic, psychological, cognitive and physiological traits. Each user having different characteristics, it is necessary to develop training procedure that are specifically adapted to each of them. This is currently not the case: training protocols are most of the time generic [6]. In order to deepen our understanding of the mechanisms underlying MT-BCI control, and consequently design training strategies adapted to each user, several studies have investigated MT-BCI performance predictors [13–15]. These predictors could explain part of the between-subject variability in terms of MT-BCI control abilities. However, for now, if a naive BCI user were unable to perform the desired mental commands due to lack of ability to produce proper EEG patterns, regardless of the signal processing that was used, we might have difficulties to explain it. Identifying predictors of MT-BCI performance could help us improve and adapt the user-training protocol to each user.

One category of factors that can be interesting to predict and explain Brain-Computer Interface (BCI) users' performances are the ones based on stable characteristic of the users such as their personality profile or age (see Chapter 5). These potential predictors are accessible through questionnaires or tests. Furthermore, as BCIs are neurotechnologies using the user brain activity to control systems, in order to better understand why some users manage to produce better quality signals (stable, strong and more discriminant) than others, we considered that neurophysiological characteristics of the users could also be relevant predictors of BCI performance (see Chapter 6). Such predictors could be accessible using a resting state baseline before the beginning of the training. In previous studies, some predictors have been identify. However, they have been revealed in single experiments, each including a small number of subjects and sessions. In addition, they correspond to univariate models and did not study how well a given predictor could actually predict the BCI performances of unseen subjects. Yet, to be useful, these models should be stable and accurate, i.e., be able to actually predict the BCI performances of new users. These models should also consider multiple variables and should generalize across experiments and datasets.

The main objective in this part is to investigate potential user-related predictors of MT-BCI performance with the view of improving MT-BCI user-training based on our results in the future. The

objective was to create computational models of BCI user training that could predict the performances of various BCI users, based on their traits, skills and their neurophysiological characteristics. Indeed, as such characteristics are accessible before the beginning of the BCI-training, being able to find a model that would predict MT-BCI performances based on those characteristics could enable us to select the (potentials) most responsive users and/or to design training protocols that are adapted to the users' profile and thereby optimises the learning abilities.

2.2.3 Contribution 3: Models based on machine-related factors

BCI identify users' intent by analyzing their brain activity, and thereby enable users to interact with the external world without any movement [16]. This activity is most often measured using EEG, and then processed using different methods at different levels: the preprocessing level, the features extraction level and finally the classification level. Therefore, ideally, when learning how to control an MT-BCI, the user also learns how to generate stable and distinct brain signals for each class (i.e. task) [**?**]. To be able to do so, extracted features from the recorded EEG signals should be consistent. In other words, they should ideally correspond to the neurophysiological patterns identified in the literature. During the calibration phase of an MT-BCI system, discriminative data-driven learning methods are commonly used to perform EEG feature extraction. Those features are then used to distinguish classes [82] using classification algorithms. One popular signal processing technique for EEG-based MT-BCIs is the Common Spatial Pattern (CSP) algorithm, which learn spatial filters that best discriminate between two MT classes [17, 18]. CSP filters maximize the variance of the spatially filtered signal under one condition while minimizing it for the other class. Before using CSP filters, several parameters have to be selected: the band pass filter, the time interval and the number of filters to use. When calibrating the system, it is common to use general settings (i.e, the same settings) for all subjects, e.g., the standard 8-30 Hz pass-band for MT-BCIs [17]. Yet, individually choosing the hyperparameters for CSP-based classification could improve online performances [19]. Some data-driven methods have been developed to select subject-specific hyperparameters [18].

Although these feature extraction and classification algorithms are commonly used and have proven to be effective, they are almost exclusively data-driven. They include very little neurophysiological prior, and rather trust the (potentially noisy) EEG data recorded during the calibration phase. In this part, we hypothesise that, to be more effective, these methods should take into consideration some constraints that may need to be identified. Indeed, are all properties of the features or classifiers learned from BCI data equally likely to be associated with good performances in practice? If not, what properties are more often associated with superior decoding performances? Would enforcing such properties in machine learning BCI algorithms lead to better decoding performances in practice?

To answer those questions, in Chapter 8, we study the impact of the properties of the most discriminant frequency band (MDFB) [18] selected by machine learning algorithms on online performances. The MDFB algorithm aims at selecting the frequency band that best discriminate between two classes. In Chapter 9 we focus on characteristics extracted from CSP algorithm and one classification algorithm frequently used in BCI: The linear discriminant analyses (LDA).

Each chapter is preceded by a review of literature. Finally, a general discussion and prospects are introduced in the last Part i.e., Part III. In addition, Figure 2.1 is a schematic representation of this roadmap which will gradually be revealed throughout the manuscript.

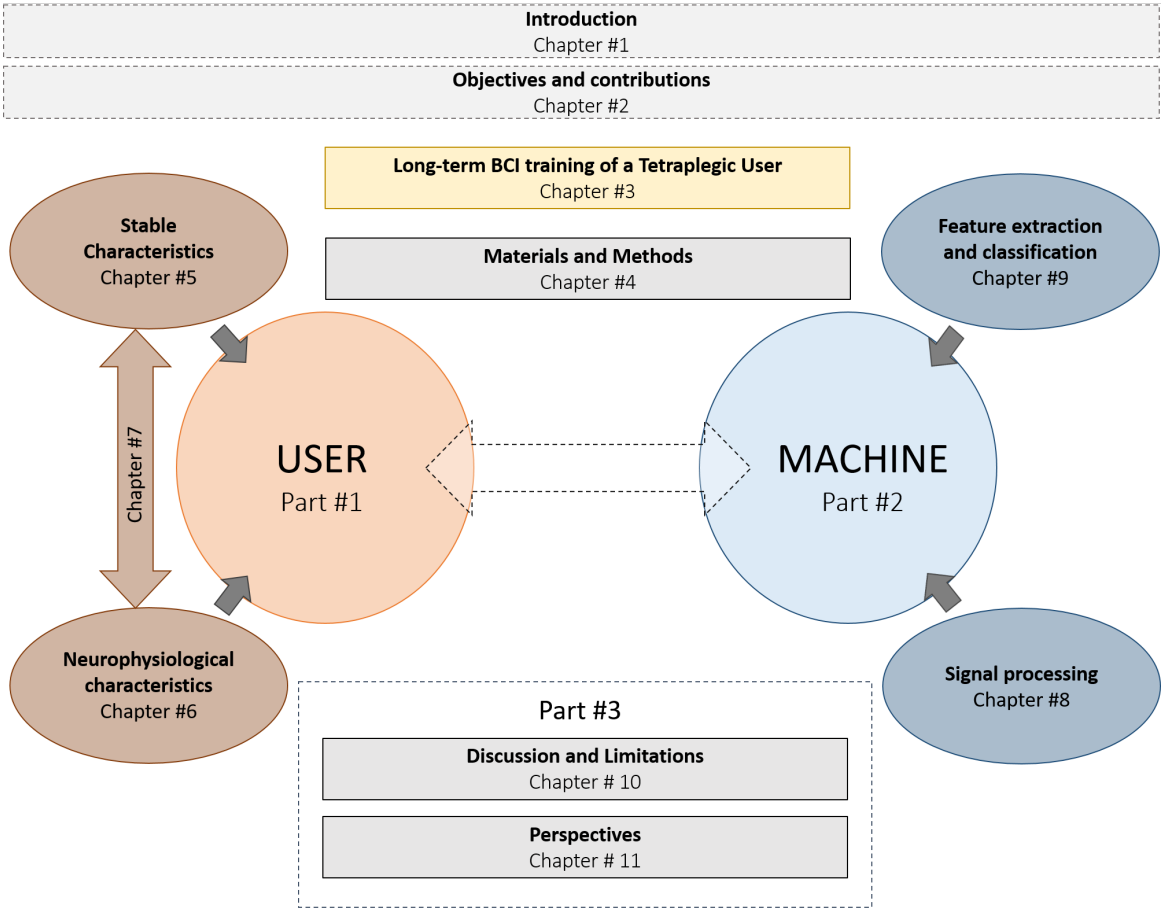
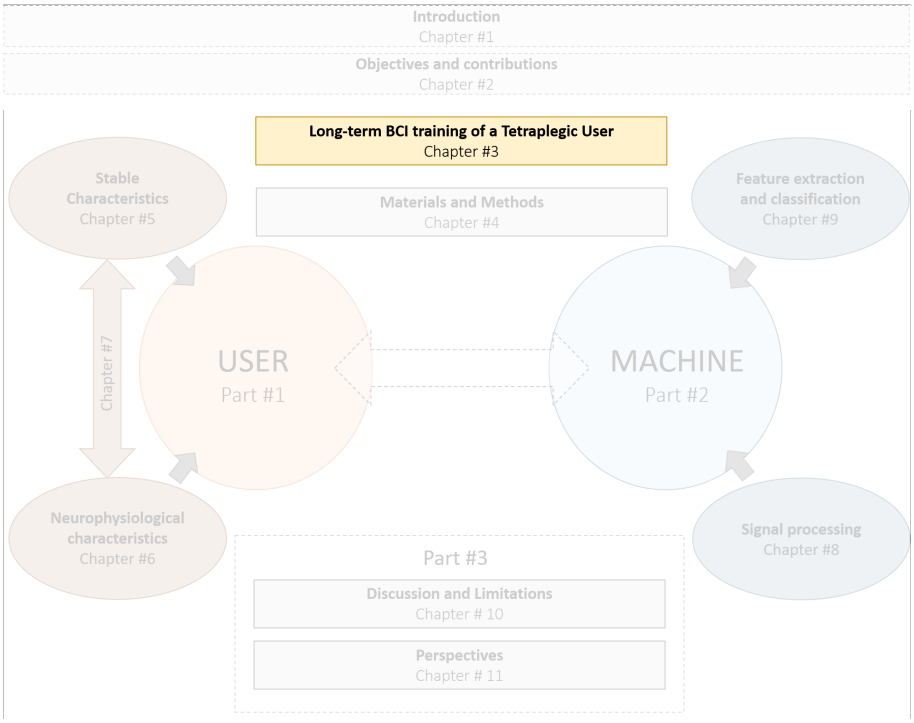


Figure 2.1: Illustration of the Roadmap of this thesis. It will be completed step by step throughout the manuscript.

3

Long-term BCI training of a Tetraplegic User

ROADMAP -



QUICK SUMMARY -

While often presented as promising assistive technologies for motor-impaired users, MT-BCI remain barely used outside laboratories due to low reliability in real-life conditions. In addition, as outlined in Chapter 2, it is necessary to understand user training in order to optimize it. Thus, in the following chapter, we propose and evaluate the design of a multi-class Mental Task (MT)-based BCI for longitudinal training (20 sessions over 3 months) of a tetraplegic user in the context of the CYBATHLON BCI series 2019. We aimed at combining a progressive user MT-BCI training with a newly designed machine learning pipeline based on adaptive Riemannian classifiers shown to be promising for real-life applications. Using a newly proposed BCI user learning metric, we could show that our user learned to improve his BCI control by producing EEG signals matching increasingly more the BCI classifier training data distribution, rather than by improving his EEG class discrimination. However, the resulting improvement was effective only on synchronous (cue-based) BCI and it did not translate into improved CYBATHLON BCI game performances. We also report on the evolution of the user's neurophysiological patterns and user experience throughout the BCI training and competition.

COLLABORATORS -

Sadatnejad Khadijeh, Roc Aline, Appriou Aurélien, Monseigne Thibaut, Pramij Smeety, Mladenovic Jelena, Pillette Léa

Contents

RELATED PAPER -

BENAROCH C., SADATNEJAD K., ROC A., APPRIOU A.,MONSEIGNE T., PRAMIJ S., MLADENOVIC J., PILLETTE L.,JEUNET C. and LOTTE F. (2021), Long-term BCI training of a tetraplegic user: Adaptive riemannian classifiers and user training", *Frontiers in Human Neuroscience* 15:118.

3.1 Introduction

This Chapter is an experimental/practical study of learning. By participating to the CYBATHLON BCI series competition in Graz in 2019, we had the opportunity to train a tetraplegic user for more than three months with a final objective, namely (winning) the competition. This experimental study was a first step in the understanding of the MT-BCI user learning before moving on to more theoretical/computational studies in subsequent chapters. Furthermore, as only a small number of studies focused on BCI end-users (e.g., users with severe motor impairment [83–85]) in real-life testing environments outside the labs [86–89], this competition gave us the opportunity to study MT-BCI technology on a end-user. The CYBATHLON was organised first by the Swiss federal institute of technology in Zurich in 2016 and then again in 2020. This event consists of races in 6 disciplines (functional electrical stimulation bike, powered arm prosthesis, powered leg prosthesis, powered exoskeleton, powered wheelchair, and BCI) with different challenges. The CYBATHLON series is a spin off of the main event that focuses on each of those disciplines. The CYBATHLON BCI series 2019, that we participated in, was held in Graz (Austria) alongside the 8th International Graz Brain-Computer Interface conference, therefore allowing the racing teams to present their technologies and methods to the whole BCI community. For this event, a computer racing game mimicking a real-life application was designed [90]. In this BCI game, tetraplegic pilots are asked to use up to four mental commands of their choice to control a virtual car.

The training of our pilot for that competition was divided into three parts in which we made adjustments in terms of learning for both the pilot and the machine. We started with an *exploratory phase*, including runs of closed-loop 2-class MT-BCI practices, where both the trainers and the trainee could apprehend the challenges, and agree upon a training routine. This was followed by a progressive training, using first a *2-class MT-BCI training phase* and then *a transfer phase* including 4-class MT-BCI training in both the standard minimalist training environment (used in previous phases) and the actual racing game environment. A progressive training seem essential for BCI as its efficiency relies on the users' ability to produce EEG patterns that are stable over time and distinct between the different mental commands [91, 92]. It is important to note that our decisions were made as we advanced in the training.

Although improving users' ability to produce such signals through user training can certainly help the participants in controlling MT-BCI [7, 12, 89], various sources of variability can lead to large shifts of data distribution between different sessions and consequently between the BCI classifier testing and training sets. Beside the largely unknown phenomena in the activity of neuronal populations which lead to non-stationarity of EEG signal [77], some variability sources including various environmental noises and changes in users' mental states such as their attention, fatigue or stress level are expected in an actual practice such as the CYBATHLON competition. When using a classifier trained on data from previous days (to avoid spending time on calibration on a new day), it tends to produce data shift between training and test sets/sessions, and thus create a BCI that is neither robust nor reliable [93, 94]. A recent signal processing approach which won multiple BCI challenges is based on the Riemannian geometry of covariance matrices [95, 96]. This approach consists in describing EEG trials through spatial covariance matrices and analyzing them in a Riemannian framework (see Section 3.2.4.2). Its unique properties led to a successful analysis of noisy data which contained many outliers [97]. Such framework does not require much training data nor the typical BCI spatial filtering data optimization to achieve high performances [96]. This motivated us to choose the Riemannian classifier for our machine learning pipeline.

Major issues in BCI signal processing include non-stationarities or signal variabilities which can be caused from e.g., changeable user skills or states [9]. These variabilities could be addressed by adaptive learning techniques in BCI [9, 94, 98]. The strength of adaptive approaches in accommodating non-stationarity led to their superiority in both online and offline BCI experiments [98]. BCI performance could also be significantly improved by combining analysis in a Riemannian framework of covariance matrices with adaptive techniques to address the omnipresence of non-stationarity [99]. Note that such work only addressed within-session variabilities, but not necessarily between session ones.

To address between-session non-stationarity effects, we propose a new method that projects all sessions to a common reference, by matching the geometric mean of a few minutes of EEG signals collected at the beginning of each session.

In this work, we address both long-term BCI user training with a BCI-naive tetraplegic user, and out-of-the-lab BCI use across multiple sessions, thus facing various non-stationarity problems. We present our approach that combines a progressive MT-BCI user training and a new adaptive Riemannian classification method that can model both within and between-session variability. Furthermore, we report on the evolution of the CYBATHLON pilot BCI classification performances, neurophysiological patterns and User eXperience (UX) along the training sessions. Regarding the study of neurophysiological patterns of user learning, beside the typical metrics, we propose a new metric to measure user learning in terms of how much the user adapts his/her EEG signals to the BCI classifier training set. Interestingly enough such metric could reveal a new form of BCI user learning. Finally, we also reflect on the pros and cons of this approach, in order to identify future areas of improvement.

This chapter is organized as follows: Section 3.2 presents the methods designed and employed for: user training, UX evaluation, EEG signal processing and machine learning and neurophysiological signal analyses. Then, Section 3.3 presents the results obtained along the training and at the competition in terms of classification performances, EEG patterns and UX. Section 3.4 discusses those results, while Section 3.5 draws lessons from them for future works. Finally, Section 3.6 concludes the chapter.

3.2 Material and Methods

As the CYBATHLON BCI series 2019 was a first experience for both the pilot and the research team, during the first 7 sessions we explored a suitable set of mental tasks and EEG sensors to use. We called this period the *exploratory phase*. Then, we progressively trained both the pilot and the machine to increase the ability of the user to generate stable and distinct brain signals for each selected task. Finally, the pilot was able to train with the racing game in which the completion time was used to evaluate performance. In the following section, we will describe our pilot as well as the protocol in each phase of the training.

3.2.1 Pilot

Our pilot, a French male who was 32 year-old at the time of the study, was injured in 2007. He has a level-C4 spinal cord injury, with an “A” score (Complete injury — No motor or sensory function is preserved in the sacral segments S4 or S5) on the ASIA impairment scale. He has a residual left shoulder motor ability.

The pilot was a naive BCI user when we started the CYBATHLON training. The study was approved by the Inria ethical committee, the COERLE (approval number 2019-12)‘, and the pilot signed an informed consent form. The training took place between June and September 2019.

3.2.2 User training

3.2.2.1 The objective: The CYBATHLON BCI race

For the CYBATHLON race, each driver sits in front of a separate screen to play the game by controlling a racing vehicle avatar. Pilots can visualize the other players on separate tracks below their own. The driver whose avatar crosses the finish line first wins the race (i.e., the race completion time is the criterion for evaluating the pilot’s performance). To control the vehicle, the pilot can modulate his EEG signals to send commands with the BCI. Depending on the course of the track, pilots can trigger three active commands (“LEFT", “RIGHT", “HEADLIGHTS"), to respectively make the car turn left, turn right or switch the headlights on, or they can just trigger the default command (“NOINPUT"), which makes the vehicle move by itself even when no input signal is sent. Figure 3.1

shows the four instructions of the game (i.e., “LEFT", “RIGHT", “HEADLIGHTS" and “NOINPUT") and their corresponding pictograms used during the user training. Indeed, before training with the game, the user trained with a classic BCI paradigm (see Figure 3.4 for the paradigm). The vehicle accelerates during the time windows where the right command is sent. A wrong command sent by accident results in a disadvantage, i.e. the avatar slows down and loses time (e.g. “HEADLIGHTS" instead of “RIGHT", or “LEFT" instead of “NOINPUT"). The method we used to decide whether or not to send a command based on the classification outputs is detailed in section 3.2.4.2.

3.2.2.2 Training schedule

Sessions typically started in the morning, taking place at the pilot's home and on a few occasions in the laboratory (for sessions 10, 12, 13, 16 and 19). Two to three researchers among our team of research were always present. The pilot training lasted approximately 3 months, between June 2019 and September 2019, and comprised a total of 20 training sessions. The training schedule was flexible, ranging from one to three training sessions a week, and from one to two hours a session. All these 20 sessions included closed-loop training runs with online BCI feedback.

The training was divided into three phases:

- **The exploratory phase**: A phase of familiarization and screening where both the trainers and the trainee could apprehend the challenge and set up a training routine (7 sessions). During this phase, several decisions had to be taken such as the Mental Tasks (MT) the pilot had to perform so that the designed BCI can send one of the four possible commands. From the first sessions, our pilot showed an interest in performing motor imagery. Therefore, motor imagery of the RIGHT-HAND and LEFT-HAND were included. Those were later associated with the controls to make the game vehicle turn right and left, respectively, to ensure a clear mapping between the MT and their effect. The rest state was included as the “NOINPUT" command and the last task (“HEADLIGHTS") had yet to be settled. As the literature suggests that a multiclass BCI benefits from selecting user-specific tasks and mixing motor and cognitive MTs [100], both MENTAL SUBTRACTION and MENTAL ROTATION were screened for the last task. MENTAL SUBTRACTION was chosen as our pilot felt more comfortable with this task. This phase enabled the pilot to apprehend the practice of Mental Tasks and to familiarize himself with the online feedback provided by the BCI classifier. During each of these sessions, he practiced in closed-loop with a 2-class MT-BCI whose classifier was trained on the data of the two first runs of that session. This phase also enabled us, the experimenters, to identify which MTs to train in a more systematic and controlled way in the subsequent phase.

- **The 2-class progressive training phase**: An intermediate phase that is part of a progressive training. Indeed, the literature on human learning suggests that it is best to train components of the task before the complete task [101]. Progressive training has previously attained high BCI performances [92]. Thus, we trained the pilot with a subset of the mental tasks before moving up to the full 4-class control. The pilot was trained to control a 2-commands BCI during 4 sessions (see Figure 3.3 for the paradigm). The training involved two different pairs of commands: two sessions were dedicated to LEFT- vs RIGHT-HAND movement imagination (7 runs in total) and two sessions to MENTAL SUBTRACTION vs REST (6 runs in total). The tasks and the number of electrodes were fixed.

- **The transfer phase**: A final training phase where the pilot alternated between 4-class online BCI and the actual CYBATHLON racing game training (hence the phase name). The game version delivered to the contestants was used for the training and then modified with random instructions order for the command, so that each race is different (only the order of the instructions during the race was changed). This phase lasted 9 sessions. The pilot started to train with the actual CYBATHLON racing game version at session 13. He was able to experience the race during 7 sessions (30 races).

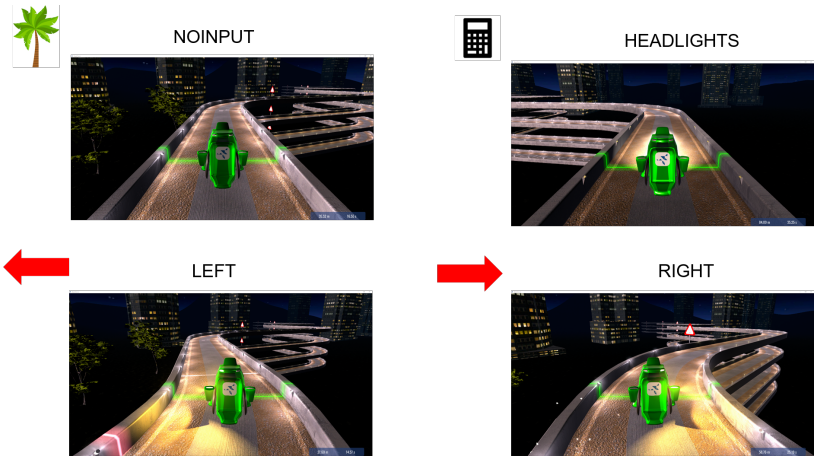


Figure 3.1: CYBATHLON racing game. We chose a palm tree as the pictogram for the REST state (i.e., the NOINPUT command), a calculator for the MENTAL SUBTRACTION task (i.e., HEADLIGHTS command), and arrows pointing to the left and right for respectively LEFT and RIGHT-HAND motor imagery (i.e., turning LEFT and RIGHT commands).

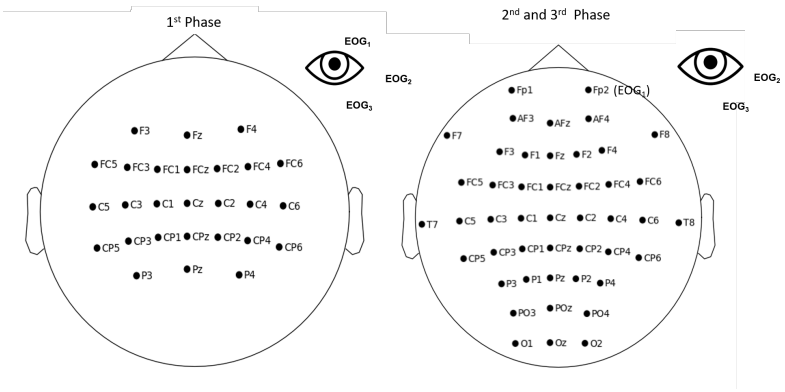


Figure 3.2: Electrodes used during the first (exploratory) phase (left), and the second (2-class training) and third (transfer) phases of the training (right). For all the phases we used the same EEG hardware g.USBAmp (g.tec, Austria), sampled at 512 Hz) but we added more electrodes(16) for the second and third phase.

At the beginning of each session, the pilot had to complete a short (about 5 min) questionnaire about his current state, and then the EEG cap was installed (about 20 minutes) while experimenters were informally chatting with the pilot. After the cap installation, two EEG baselines (resting state) were recorded respectively with eyes open and eyes closed (2*3min). To record the EEG signals, a different number of active scalp electrodes, referenced to the left ear, were used in the three training phases (see Figure 3.2). EEG signals were measured using a g.USBAmp (g.tec, Austria), sampled at 512 Hz and processed online using OpenViBE 2.2.0. [102]. We decided to increase the number of electrodes between the exploratory and 2-class training phase. Contrary to the setup used in the exploratory phase that only used motor-related electrodes, the additional electrodes served mainly for MENTAL SUBTRACTION and REST MT.

For the exploratory phase, we used the standard "Graz BCI" bar feedback [103], as implemented in OpenViBE for 2-class MT-BCIs and as used in, e.g., [104, 105]. In that phase each run included 20 trials for each of the two MT classes. A classifier was built on the data from the two first runs to provide online feedback for the subsequent training runs of that session. Due to the exploratory and non-systematic nature of that phase, the number of runs performed fluctuated largely between sessions.

The same paradigm was used for all online BCI runs during the 2-class training phase (see Figure 3.3). Each run comprised 40 trials (i.e., 20 trials per class). The number of runs depended on the session duration and the pilot's state, notably his motivation and fatigue. For each trial, a cross

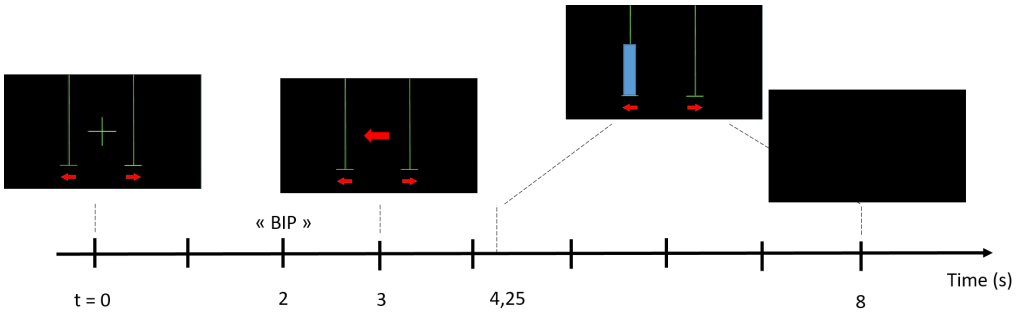


Figure 3.3: Experimental paradigm used for the 2 classes MT-BCI and more specifically here for the 2-class motor imagery training (LEFT-HAND vs RIGHT-HAND motor imagery).

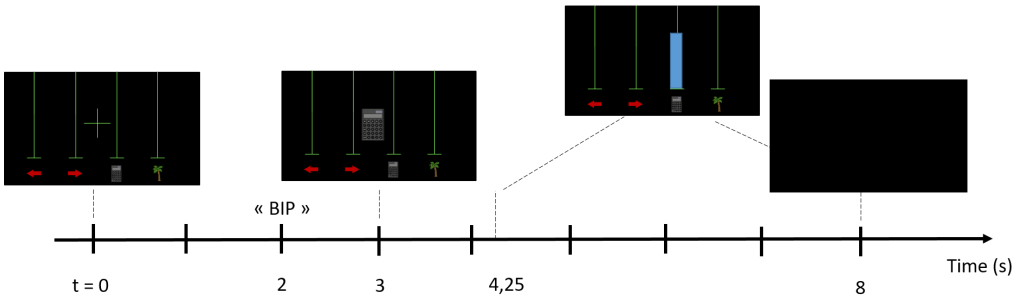


Figure 3.4: Experimental paradigm used for the 4 classes MT-BCI

was first displayed. The Mental Task (MT) to be performed was then announced by a “beep" and the corresponding pictogram appeared in the middle of the screen, i.e. an arrow pointing to the left representing a LEFT-HAND motor imagery task; an arrow pointing to the right representing a RIGHT-HAND motor imagery task; a calculator representing a MENTAL SUBTRACTION task; a palm tree representing the REST state, chosen together with the pilot. Then, a blue bar was displayed as continuous visual feedback. The location of this bar (i.e. over which pictogram it appears) indicated the MT recognized by the classifier and its length the classifier confidence in this recognition. The bar was displayed only when there was a match between the instruction and the recognized task, i.e., it was a positive only feedback. During the training phase, we alternated between 2-class and 4-class MT. Hence, we used the same paradigm for the 4-class phase (Figure 3.4). Each run comprised 40 trials (i.e., 10 trials per class) but this time the pilot had to perform one of the four chosen tasks for the game.

After the cap removal, the majority of the sessions ended with a quick end-of-session questionnaire on the computer (about 5min) and an informal debriefing concerning the mental strategy used to perform the tasks and, in response to the pilot’s inquiries, the performances achieved, the meaning of the feedback or the classification method used.

3.2.3 User experience

3.2.3.1 Mental Tasks

The four tasks that were used are those described in Figure 3.1, they were chosen according to the pilot’s wishes and a short screening. Regarding the Motor Imagery (MI) tasks associated with paralyzed limbs (i.e. Left-Hand (LH) and Righ-Hand (RH)), the pilot experimented with several options to determine what he felt most comfortable with (e.g., imagination of opening and closing his hand) and finally settled on the imagination of boxing moves for each hand. The third Mental Task (MT), i.e. MENTAL-SUBTRACTION (MS) task, was initially conducted with cues in the exploratory phase, where the pilot was shown 3-digit numbers on the screen and was instructed to gradually subtract randomly generated two-digit numbers from them. Then, in the 2-class BCI

training phase and the transfer phase, the pilot could perform mental math without cue, i.e. he was instructed to spontaneously choose a “random" number and to make the corresponding subtractions of 2-digit numbers by 1-digit numbers. He felt comfortable with this third task.

3.2.3.2 Questionnaires

Before and after most of the sessions, our pilot completed questionnaires regarding his subjective states (see Appendix C in [106]). Both questionnaires retrieved user experience (UX-) related factors based on subjective 5-point Likert scale in 8 items (pre-session) and 21 items (post-session). Based on these items, a score can be computed for 5 factors, i.e. mood, motivation and mindfulness (assessed pre- and post-session) along with post-session agency (feeling of control over the feedback provided by the system) and cognitive load (amount of cognitive process required to control the system).

3.2.3.3 Interview

After the competition, an interview was conducted with the pilot. The aim was to analyse qualitatively the impact that the whole process (from the first contact, to the BCI training and competition) had on the pilot's representations related to BCIs. More precisely, using a semi-structured interview approach, we investigated the pilot's initial acceptability of BCIs, including different dimensions such as motivation, perceived usefulness, perceived ease of use and intention to use BCI-based technologies. Then, we sought to understand how his CYBATHLON experience modified these representations, and what level of acceptance resulted from this “adventure". Due to the COVID-19 crisis, the interview was conducted remotely (video conference). It lasted one hour and started with an explanation of the objectives and structure of the conversation to come. It was divided into four parts dedicated to different moments: i) before the BCI training, ii) during the BCI training, iii) during the CYBATHLON competition and iv) after the CYBATHLON competition. Each part was divided into a series of questions related to different dimensions of acceptability and acceptance, mainly based on the Technology-Acceptance Model (TAM) 3 [107] questionnaire. The TAM3 suggests that one's usage behaviour is determined by their usage intention, itself being influenced by both the perceived usefulness and perceived ease of use of the technology. According to the TAM3, the perceived usefulness and ease of use would be influenced by social norms, inter-individual differences (i.e., psychological traits and states and socio-demographic characteristics), facilitating conditions and characteristics related to the technology. The influence of these factors would be modulated by the fact that the technology use is voluntary or not, and by the experience that the user has with the technology.

3.2.4 Signal processing and machine learning

Due to the exploratory, and non formal nature of this work (a competition preparation), the machine learning tools we used evolved along with the training, according to the problems we encountered. During the exploratory phase, we first started with a standard MT classification pipeline based on Common Spatial Pattern (CSP) spatial filtering [17] and a Linear Discriminant Analysis (LDA) classifier [108], for the exploratory phase. For this phase, we calibrated the CSP and LDA on the first two runs (acquisition runs) of each session. Due to sensitivity of CSP to noise and outliers [109, 110], and since we performed user training mostly at home, an environment with different variability sources, a more robust approach was used for the next phase. During the 2-class training phase, we thus represented EEG by spatial covariance matrices and analysed them in a Riemannian framework [95, 96]. Indeed, such Riemannian classifier proved very efficient for EEG signal classification in several offline EEG classification competitions [95, 96]. In this phase, we also calibrated the Riemannian classifier on each session acquisition runs.

For the competition day, we needed a previously trained classifier to avoid calibration time. Moreover, recalibrating the classifier everyday could lead to an ever-changing feedback which may be detrimental to user training [12, 88, 89]. Therefore, for the transfer phase (session 12 and subsequent)

we calibrated our classifier based on the runs from the previous phase sessions that were the least noisy and the least contaminated by artifacts. Nonetheless, for the two first sessions of the transfer phase, considerable shifts and performance variation between sessions were observed. Therefore, we started using an adaptive approach from session 14 onward. This approach consisted in reducing non-stationarity notably by projecting all training sessions data into a common subspace, and then in adaptively projecting the test set to this common subspace. We describe the technical details of the preprocessing, features, classifiers, adaptation method, inter-session variability reduction, and evaluation criteria in this section.

3.2.4.1 Preprocessing

During the first two phases where the classifier was trained on the data of dedicated in-session runs, EEG signals were processed by segmenting them into 1 second windows starting from 1.25 second after the start of the instruction cue, with 93.75% overlapping (i.e. 1/16s shift) both for training the classifier and online-test. During the last phase (i.e. transfer), since we started using multiple training sessions, different overlaps were used for the classifier calibration and for the online classification. For the calibration of the classifier, notably for computational and memory reasons, we chose smaller overlaps between consecutive windows (between 50% to 87.5% depending on the classifier). For online evaluation, since a continuous feedback was needed, we used 93.75% overlapping between consecutive windows for session 12 and 13 and 87.5% from session 14 and for the game. All EEG signals were band-pass filtered in 8-24 Hz using a butterworth filter of order 4.

Under the CYBATHLON BCI race regulation, some preprocessing was applied to ensure that the pilot controls the avatar using signals originating solely from brain activity and not signals of a muscular origin. The main goal of this preprocessing was rejecting artifacts. We rejected all EEG epochs in which band-pass filtered EEG had absolute amplitude higher than 70 μV, or included horizontal or vertical eye movements. To detect eye-movement related artifact, three electrodes were put around the left eye, i.e. EOG1, EOG2, and EOG3. Vertical eye movements were detected with EOG1 and EOG3 put on the vertical line of the eye, up and down respectively. EOG2 was put on the horizontal line of the eye. The electrodes placement is shown in Figure 3.2. The EOG signals were filtered in 1-10 Hz using 4-order butterworth filter.

To detect the possible eye artifacts, we computed two vertical and horizontal signals as follows:

$$V = EOG1 - EOG3 \tag{3.1}$$

$$H = EOG2 - \frac{(EOG1 + EOG3)}{2} \tag{3.2}$$

where V and H are related to vertical and horizontal eye movements respectively. After preprocessing, each EEG window that did not include a local peak of vertical or horizontal EOG was considered as clean and used for model training and online classification. Otherwise, it was rejected. In order to find the local peaks of these signals, thresholds were applied over the average absolute value of the vertical/horizontal signals in each window as follows:

$$\frac{1}{T} \times \sum_{t=1}^{t=T} |V_t| < T_V \tag{3.3}$$

$$\frac{1}{T} \times \sum_{t=1}^{t=T} |H_t| < T_H \tag{3.4}$$

where T_V and T_H denote the thresholds for finding vertical and horizontal eye movement, T denotes the time length of each window, V_t and H_t denotes the vertical and horizontal signal at t^{th} time point of each window. The threshold values for vertical and horizontal signals were estimated on training data as follows:

$$T_S = \frac{\sum_{t=1}^{t=N} |S|}{N} + 3\sigma_{|S|} \tag{3.5}$$

where S could be either V or H signals, $\sigma_{|S|}$ is the standard deviation of the signal $|S|$, and N denotes the time-length of the signal.

3.2.4.2 Features and Classifiers

Due to the exploratory nature of this work we switched between two main online classification pipelines for MT-BCI between the exploratory and the 2-class user training phases. The details of the features and classifiers used online during the different training phases are as follows:

CSP+LDA:

In the exploratory phase, we used standard linear classifiers and spatial filters, to combine several channels (here with a linear combination) into a single one [17, 111].We notably used the Common Spatial Pattern (CSP), a popular approach that extract discriminative spatial filters for classification [17, 111]. In our experiments, for the exploratory phase, we used 6 CSP filters, corresponding to the 3 largest and 3 smallest eigenvalues [17, 111]. We then used Linear discriminant Analysis (LDA), a linear discriminative classifier, for classifying CSP-based features [108].

Riemannian classifiers:

Due to the presence of different noise and artifact sources during the exploratory phase, and due to the sensitivity of CSP to them, we moved on to analysis in a Riemannian framework for the 2-class training phase and continued with it for the transfer phase. EEG analysis in a Riemannian framework consists in describing EEG epochs by their Spatial Covariance Matrices (SCM) and using Riemannian Geometry to consider the non-linear geometry of the space in analyses. In this section, we first define the principles of analysis in a Riemannian framework, then we describe the Riemannian classifiers we used. For describing the EEG epochs as SCMs, we used the optimal linear shrinkage estimator of SCM from [112]. We denote each SCM by $C_j^{(c_i)}$, where c_i is its corresponding class and j its epoch index. $C_j^{(c_i)}$ is an $N \times N$ Symmetric Positive Definite (SPD) matrix, where N denotes the number of EEG channels. The space of SPD matrices, which is not closed under scalar multiplication, is not a vector space. We reformulate the feature space as a Riemannian manifold by equipping each tangent space with a Riemannian metric. The Riemannian distance, which is the distance between two points (SCMs) along the manifold, can be computed as follows [95]:

$$d_R(C_i, C_j) = \|log(C_i^{-1/2} \times C_j \times C_i^{-1/2})\|_F = (\sum_{l=1}^{N} log^2(\lambda_l))^{1/2} \tag{3.6}$$

where i and j are the indexes of the epochs, $\|.\|_F$ denotes the Frobenius norm, and λ_l is a positive eigenvalue of the $C_i^{-1/2} C_j C_i^{-1/2}$ matrix. By changing the analysis framework to a Riemannian framework, we needed to adapt geometrical and statistical concepts to this geometry. In the following, we only describe some definitions that we need for this chapter

Riemannian Mean: The mean points of SPD matrices, which is called Karcher/Frechet mean, is defined as a point over the manifold (i.e., an SPD matrix) which has minimum squared Riemannian distance from all the points over the manifold.

$$\overline{C}^{(c)} = argmin_C \sum_{i=1}^{N} d_R^2(C, C_i^{(c)}) \tag{3.7}$$

where $\overline{C}^{(c)}$ denotes the geometric mean of N samples from class c [113]. Since there is no closed form solution to this problem, we used the iterative estimator proposed in [114].

Riemannian variance: The sample variance over the manifold, which is known as Frechet variance, represents the dispersion around the mean point over the manifold. We compute the sample variance as the expected value of the squared Riemannian distance around the geometric mean [113].

$$\sigma^{(c)^2} = \frac{1}{N_c} \times \sum_{i=1}^{N_c} d_R^2(C_i^{(c)}, \overline{C}^{(c)}) \tag{3.8}$$

Fisher geodesic Minimum Distance to Mean (FgMDM): FgMDM is a pipeline combining Fisher Geodesic Discriminant Analysis (FGDA) and Minimum Distance to Mean (MDM) classifier [115]. Geodesic filtering is achieved in tangent space with a Linear Discriminant Analysis to keep class-related information and discard irrelevant information. After filtering, data are projected back to the manifold and classified using a Minimum Distance to Means (MDM) classifier [115]. An MDM classifier works as follows: for training, it computes the Karcher mean of the training SCMs from each class. Then, for predicting the label of a test SCM, MDM computes the Riemannian distance of this SCM to each class mean and assigns it the label of the nearest class mean. The details of the FgMDM algorithm for training and testing can be seen in algorithms 3 and 4 of Appendix B respectively.

Adaptive rebias FgMDM: As mentioned before, in order to reduce both within-session variability and shifts between training and test sets, we started using an adaptive approach in the transfer phase, from Session 14 onward. To do so, we used an adaptive rebias FgMDM classifier [99, 116]. Rebias FgMDM relies on aligning the covariance matrices from the training and test set onto a common reference. For this purpose, the idea is to identify a projector, one for the training set and one for the testing set, so that the Karcher mean of the projected SCMs for each set ends up on a common reference, here the identity matrix. With this approach, all SCMs C_i' are thus centered around the identity matrix, using the following transformation:

$$C_i' = \overline{C}^{-1/2} \times C_i \times \overline{C}^{-1/2} \tag{3.9}$$

where $\overline{C}$ denotes the Karcher mean of the training/test set for projecting the training and test SCMs respectively. Note that this projection does not require class labels, and is thus unsupervised. It should also be noted that when used online, the test set is not fully available, and thus the real $\overline{C}$ cannot be estimated. An adaptive estimation of it, based on incoming test SCMs, is thus used. More precisely, the test set Karcher mean is adaptively estimated using a weighted geodesic interpolation between previous estimates and each ongoing SCM, as follows:

$$\overline{C}_{i+1} = \overline{C}_i^{1/2} \times \left(\overline{C}_i^{-1/2} \times C_{i+1} \times \overline{C}_i^{-1/2}\right)^{1/(i+1)} \times \overline{C}_i^{1/2} \tag{3.10}$$

where $\overline{C}_i$ is the current estimates of the test set Karcher mean, C_{i+1} denotes the ongoing SCM epoch, and $\overline{C}_{i+1}$ denotes the updated test set Karcher mean. $\overline{C}_{i+1}$ is then used to project the subsequent test SCMs using Eq. 3.9. This estimate is initialized to the training set Karcher mean. See [99] for more details on this adaptive rebias FgMDM. To the best of our knowledge, this is the first time that such an adaptive Riemannian classifier is used online for MT-BCI.

Whereas adaptive rebias FgMDM could reduce the shifts between the training and test sets as well as within-session variability, it does not address shifts and variability that may occur between sessions. However, our classifier is trained on multiple sessions, and aims at being applied unchanged on multiple sessions as well. We thus needed to address between-session variability as well.

Reducing between-session variability: From the beginning of the transfer phase, we started using a classifier trained over multiple runs from multiple sessions. However, we observed considerable shift between different sessions, resulting from different non-stationarity sources. Figure 3.5 illustrates the two dimensional representation of two first runs of sessions 12-18. All these sessions were 4-class training sessions (LEFT-HAND vs RIGHT-HAND vs MENTAL-SUBTRACTION vs REST) recorded

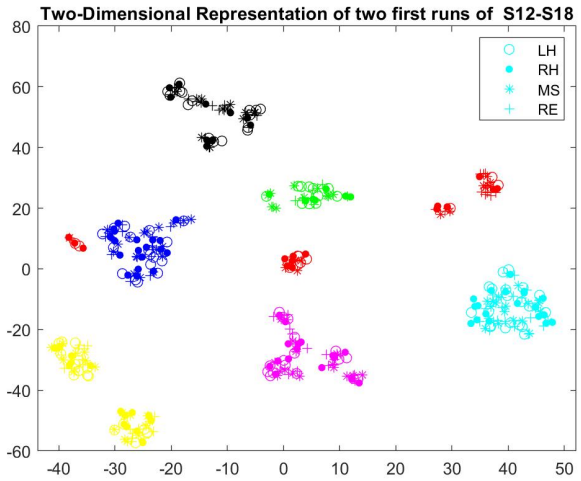


Figure 3.5: Two-dimensional representation of SCMs, projected by t-SNE, recorded in two first runs of sessions 12-18 (different sessions are illustrated in different colors, S12:green, S13: blue , S14: black, S15:magenta, S16: yellow, S17: red, S18: cyan).

during the transfer phase. We used t-SNE [117] to project the data points (EEG covariance of one epoch/trial) in two-dimensional space. For dimensionality reduction using t-SNE to consider the non-linear geometry of the feature space, we used the Riemannian distance as the custom distance parameter. In addition, we set the effective number of local neighbors of each point (i.e., the Perplexity parameter) to the number of epochs extracted from each task in each run (i.e., 10).

As mentioned before, adaptive rebias FgMDM [99] only addresses within-session variabilities, but not between-session ones. Thus, to reduce the latter we explored a new approach, which aimed at projecting all sessions (both training and testing ones) to a common reference. More precisely, we estimated a projector for each session, whose purpose was to project the Karcher mean of the two-minutes baseline recorded at the beginning of that session (during which the pilot was in a resting with eyes open condition) to the common reference, i.e., the identity matrix. Then, all MT-BCI SCMs of that session were projected using this session-specific projector, as follows:

$$C'_{i,j} = \overline{R_j}^{-1/2} \times C_{i,j} \times \overline{R_j}^{-1/2} \tag{3.11}$$

where $\overline{R_j}$ is the Karcher mean of epochs extracted from the baseline recorded at the start of each session, j denotes the session index and $C_{i,j}$ is the i^{th} MT-BCI SCM sample of the j^{th} session. Such projected SCMs were then used as input to the adaptive rebias FgMDM classifier described above. The final classifier was thus trained on multiple sessions that were all projected to the same common reference, whereas the online data were also projected to this common reference, thanks to the projector built on the baseline of that session.

This approach was used for all sessions from Session 14, including during the day of the CYBATHLON competition. To the best of our knowledge, this new approach is the only one aimed at reducing between-session variability with Riemannian classifiers, in a completely unsupervised way, and without requiring MT-BCI data (but only a baseline). Moreover, such approach could also be used with non-Riemannian classifiers, e.g., with classical CSP-based BCIs, since CSP extracts features from covariance matrices as well. It should be mentioned that, within the BCI community, it is still debated whether adaptive machine learning algorithms do favor user learning or not, and if so, how and how often the adaptation should be performed (see [12, 89, 118] for reviews). In this study, we tried to favor a beneficial mutual adaptation between the user and the machine by using a machine learning adaptation that was purely unsupervised and class-unspecific. Indeed, our adaptive Riemannian classifiers only tracked and adapted to the global EEG changes, but not to the

class-specific EEG ones. In particular, it should be stressed that our FgMDM classifier weights W (see Appendix B) were not changed by the adaptation - only the projections applied to the input covariance matrices (Eq. 3.9 and 3.11) were adapted. This way we hoped to address global EEG non-stationarity while limiting the risk of an unstable feedback that would happen if constantly updating the classifier weights.

Postprocessing:

For the standard bar-feedback MT-BCI training (illustrated in Fig. 3.4), the feedback was continuous and directly related to the classifier output. Here the FgMDM classifier outputs (Riemannian distances to class means) were transformed into pseudo-class probabilities P_i, as $P_i = \frac{p_i}{\sum(p_i)}$, with $p_i = \frac{min_i d_i}{d_i}$, $d_i = d_R^2(C, \overline{C_i})$, and C being the current EEG epoch covariance matrix (after projection and geodesic filtering). The feedback bar length was directly proportional to P_i.

For the CYBATHLON game, the commands sent and the resulting feedback provided were not continuous but discrete, e.g., the car headlights were either on or off. Since sending erroneous commands was slowing the car, it was also important to reduce false positives, and thus to send commands to the game only when the BCI was confident enough that this was the correct command. Thus, for the CYBATHLON game, the BCI sent a command to the game only when the classifier identified the same class label consecutively for the last N_e EEG epochs, and with an output pseudo-probability greater than P_c. In practice, we empirically determined suitable values for our pilot, and used (after a few trials and errors) $N_e = 8$, and $P_c = 0.3$ (for a 4-class problem, the minimum probability to select a class was thus 0.25). As a reminder, for the game we used sliding EEG epochs that were 1 second long, with 1/8 second step between consecutive epochs.

Evaluation criteria:

To evaluate the user's progress, besides classification accuracy, we also used studied two additional metrics that study users' EEG changes: *classDis* and *Test-Train Adaptation*.

classDis measures how distinct and stable the EEG patterns produced by the user are, independently of any classifier [119]. For a two-class problem, *classDis* is defined as follows:

$$classDis(c_i, c_j) = \frac{d_R(\overline{C}^{(c_i)}, \overline{C}^{(c_j)})}{\frac{1}{2}(\sigma_{c_i} + \sigma_{c_j})} \tag{3.12}$$

where the numerator computes the distance between class means and the denominator is the summation of average squared distances around the Karcher mean of each class. For multi-class problems, the numerator of *classDis* computes the distance of the Karcher mean of each class $\overline{C}^{(c_i)}$ from the Karcher mean of all the data $\overline{C}$, as follows:

$$classDis(c_i) = \frac{\sum_i d_R(\overline{C}^{(c_i)}, \overline{C})}{\sum_i \sigma_{c_i}} \tag{3.13}$$

where $\overline{C}^{(c_i)}$ and $\overline{C}$ are the Karcher mean of each class and of all the data respectively. Overall, *classDis* could thus measure whether the user's EEG patterns for each class become increasingly more distinct with training, independently of any classifier.

However, *classDis* may not capture all forms of BCI user learning. In particular, since *classDis* is classifier-independent, it may not capture some potential user adaptation to the BCI system and classifier. Indeed, with BCI training, some users may learn to produce EEG patterns that become increasingly more systematically similar to those expected by the classifier, i.e., to those used to train the classifier. In turn, this would lead these EEG patterns to be increasingly more correctly recognized, and hence the user to reach increasingly better MT-BCI control. Thus, in order to evaluate whether some user learning occurred as users' EEG signals changing to adapt to the BCI

system/classifier, we need a new metric. In particular, we need to quantify how much the users' online EEG data, i.e., test data distribution, becomes similar to the EEG data used to train the classifier, i.e., training data distribution. We propose here a new criterion in the Riemannian framework of covariance matrices to do so.

To estimate the similarity/dissimilarity between training and test sets, we chose multiple landmarks from the training and test sets: we used the Riemannian means of each class as well as the overall Riemannian mean of all samples (all classes together). The smaller the distances between these landmarks in the training and test set, the more similar their distributions. Here, the distance between these landmarks between the training and test sets are normalized by either the within-class variance or the overall variance in the training set. The average of these normalised distances represent the similarity between the test and training sets, and can thus be used to measure test-train adaptation (TTA):

$$TTA = \frac{1}{(N_c+1)} \left(\sum_{i=1}^{i=N_c} d_R(\overline{C}_{train}^{(c_i)}, \overline{C}_{test}^{(c_i)}) / \sigma_{train}^{(c_i)} + d_R(\overline{C}_{train}, \overline{C}_{test}) / \sigma_{train} \right) \tag{3.14}$$

where $\overline{C}_{train}^{(c_i)}$ and $\overline{C}_{test}^{(c_i)}$ denote the Riemannian mean of class c_i in training and test sets respectively. $\sigma_{train}^{(c_i)}$ denotes the standard deviation of class c_i in training set and $\overline{C}_{train}$ denotes the global mean of all training samples. Here, all covariance matrices were recentered using Eq. 3.9 and 3.11, as done online to reduce between- and within-session non-stationarities. N_c is the number of classes. This TTA metrics displays some similarities with the "classifier precision" metrics from [120]. Indeed, this latter metrics quantifies the similarity between test trials and the feature class distribution of a classifier. It was used in [120] to assess how much an adaptive classifier class distribution could adapt to online testing EEG data. Here we use TTA in the opposite way, to assess how much the user's testing EEG data adapt to the classifier class distribution. Overall, a decreasing *TTA* metric over online sessions would suggest more adaptation of test sets to the training set. In other words, this would mean that the user is producing EEG patterns increasingly more similar to what the classifier expects for each class.

3.2.5 Neurophysiological analysis

For all our neurophysiological analyses, we rejected artifacted epochs (i.e., artifacted trials) from our data, i.e., epochs with absolute amplitude higher than 70 μV. Neurophysiological analyses were made for the 2-class BCI training phase and the transfer phase (i.e., 2-class online BCI and 4-class online BCI). We used a robust variant of the Fisher score (FS) to evaluate the discriminability between the different classes for different frequency bands and brain areas:

$$FS = \frac{\sum_{i=1}^{N} (m_i - m_0)^2}{\sum_{i=1}^{N} mad_i^2} \tag{3.15}$$

where $m_0 = \sum_{i=1}^{n}(m_i)$, m_i is the medians and mad_i the Median Absolute Deviations (MAD) of the EEG signal power spectral density distributions for the i^{th} mental class in a specific frequency band and channel, within each run. N is the number of classes. This FS formulation uses a median and a MAD instead of a mean and a standard deviation in the usual FS, the former estimators being more robust estimators than the later ones, for the expected value and dispersion around the expected value, respectively.

In our analyses, we considered three different frequency bands: the α-band (8-12 Hz), the low β-band (13-20 Hz) and the high β-band (21-30Hz), since all these bands are known to be involved in cognitive and motor imagery tasks [100]. We chose not to analyse the sessions of the exploratory phase since both the setup and the user training protocol were not fixed yet. During the second user training phase, the pilot was trained with two pairs of tasks, including training to control a 2-class *motor imagery* BCI (imagination of a left vs right hand movement). Thus, we computed the mean of the FSs obtained over the electrodes of the motor area [103] (see Figure 3.6) for each frequency band to evaluate in which frequency band the discriminability between the 2 classes was the highest.

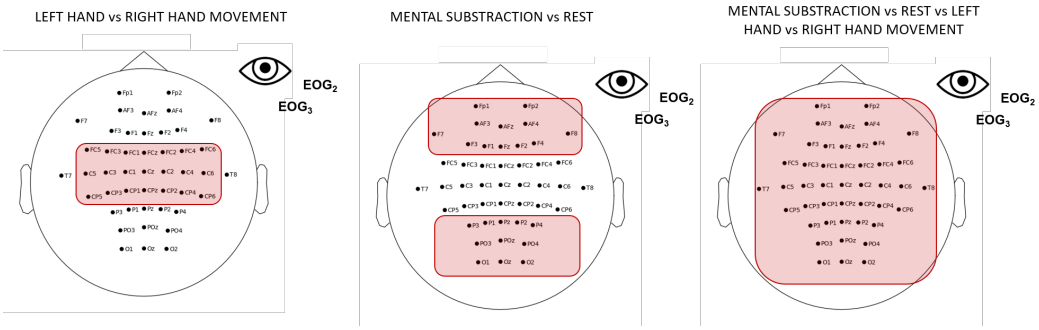


Figure 3.6: Selected electrodes for neurophysiological analysis. From left to right: Electrodes of the motor area for the LEFT- vs RIGHT-HAND motor imagery tasks, electrodes of the frontal, parietal and occipital areas for the MENTAL SUBTRACTION vs REST mental tasks and all the electrodes for the 4-class BCI.

During this second phase, the pilot was also trained to control a 2-class BCI with *cognitive* tasks (REST state vs MENTAL SUBTRACTION). As for the motor imagery tasks, we computed the mean of the FSs obtained for each frequency band. This time, we selected electrodes of the frontal, parietal and occipital areas (see Figure 3.6). We chose to focus on those specific areas because MENTAL SUBTRACTION can involve frontal as well as parietal processes [121].

Finally, during the transfer phase in which it was a 4-class *mental tasks* BCI, we did the same analysis with all the electrodes (see Figure 3.6).

We assessed possible learning effects by computing the Pearson correlation between the FS and the run index, as in e.g., [88], for each frequency band and each phase of the training.

3.3 Results

3.3.1 Behavioral results

To evaluate the learning progress across different sessions and within each session, we studied both the BCI performance in terms of online classification accuracy (Figure 3.7) and offline *classDis* (Figure 3.8) for feedback training sessions. We computed *classDis* for all runs, including the acquisition runs (for sessions that had such runs) and the test runs (Figure 3.8). For game performance we evaluated the user performance in terms of required time to complete the game.

3.3.1.1 Exploratory phase

Due to the exploratory and non-systematic nature of this phase, online classification accuracies were not recorded for every run or session. The mean online accuracy for the runs where this data was recorded is reported in Figure 3.7.(A). The dashed line in this figure shows the upper-confidence limit of the chance level for $\alpha = 0.05$ and 20 trials/class [122]. The non-systematic ways the sessions of that phase were performed also made comparing *classDis* across sessions non-meaningful. We thus did not report it for that phase.

3.3.1.2 2-class training phase

For the two-class training phase, two different pairs of tasks were used, with different discriminablity. Thus, we report the results related to MENTAL-SUBTRACTION vs REST tasks imaginations (Session 8 and 10) in a sub-plot (B) and the results related to session 9 and 11 (LEFT-HAND vs RIGHT-HAND motor imagery experiments) in sub-plot (C) in Figures 3.7 and 3.8. For the 2-class training phase, the upper confidence limit of the chance-level for classification accuracy was computed for a significance level of $\alpha = 0.05$ and for 20 trials/class [122]. We can observe that the performance with MENTAL-SUBTRACTION vs REST is clearly superior to that with LEFT-HAND vs RIGHT-HAND. The learning progress in terms of the Pearson correlation of the run classDis with

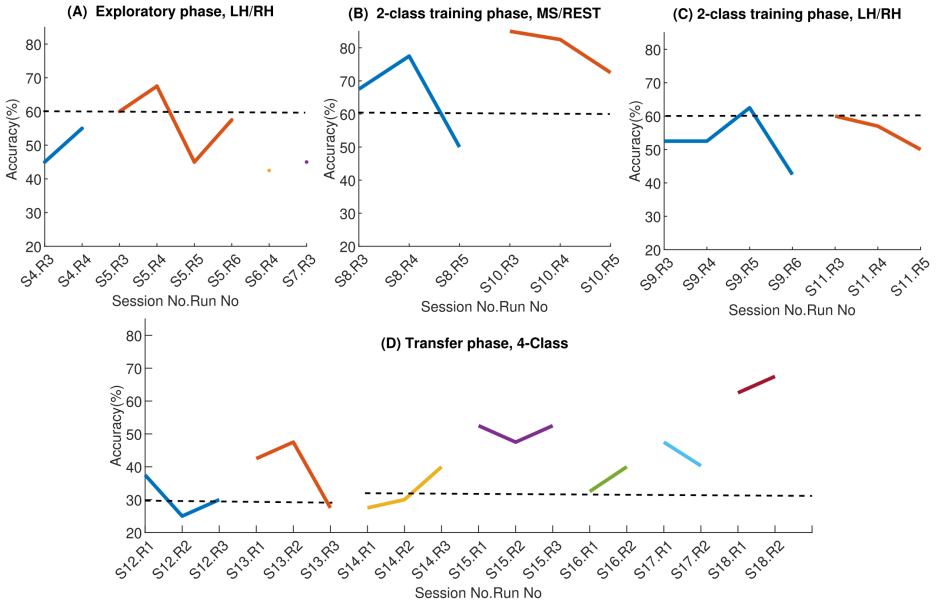


Figure 3.7: Online classification accuracy across sessions and runs **(A)** Exploratory phase, **(B)** MENTAL-SUBTRACTION vs REST of two-class training phase, **(C)** LEFT-HAND vs RIGHT-HAND of two-class training phase, and **(D)** Transfer phase. Each point represent the accuracy of a run, different colours represent different sessions. The solid lines show the online accuracy and the dashed lines show the upper confidence limit of the chance-level at $\alpha = 0.05$, according to muller2008better.

the run index in different frequency sub-bands did not reveal a significant learning for any band nor any class pair.

3.3.1.3 Transfer phase

The sessions of the transfer phase include both 4-class feedback training and game training. The results of the feedback runs in terms of classification accuracy and *classDis* are reported in sub-plots (D) and (C) of Figures 3.7 and 3.8 respectively. In Figure 3.7.(D) the upper confidence limit of a random classification accuracy in the transfer phase, illustrated using a dashed line, was computed for a significance level of $\alpha = 0.05$ and for 10 trials/class [122]. In terms of classification accuracy, a very clear performance improvement between sessions can be observed. Indeed, performances for the 4-class BCI started at near chance level (between 25% to 37.5% for the first 4-class session) and finished at 67.8% for the last run of the last session, while generally increasing (with fluctuations) in between. A Pearson correlation between classification accuracy and run index revealed that this learning effect is significant ($\rho = 0.65$, $p < 0.01$). Comparing 4-class feedback training classification accuracy before (Sessions 12 and 13) and after using the proposed adaptive approach (Sessions 14 to 18) provides interesting insights. In terms of run accuracy, a t-test revealed a trend towards run accuracy being higher with the adaptive approaches rather than without ($p = 0.07$). More interestingly, we also estimated the learning progress using the Pearson correlation of the run accuracy with the run index. Such analyses revealed no learning effect before using the adaptive approach ($\rho = 0.18$, $p = 0.73$), but a clear learning effect after the user started to use the adaptive approach ($\rho = 0.68$, $p = 0.015$), with a mean accuracy of 44.5% for the two first sessions and 54.45% for the last two sessions. This thus suggests that our proposed adaptive approach may have contributed to improve user training, by providing a more stable classifier, and therefore a more stable feedback.

On the other hand, offline *classDis* did not show any significant correlation with run index in any band.

This apparent inconsistency between significant improvement of online classification accuracy but non-significant improvement of *classDis* in the transfer phase, could be explained by studying our

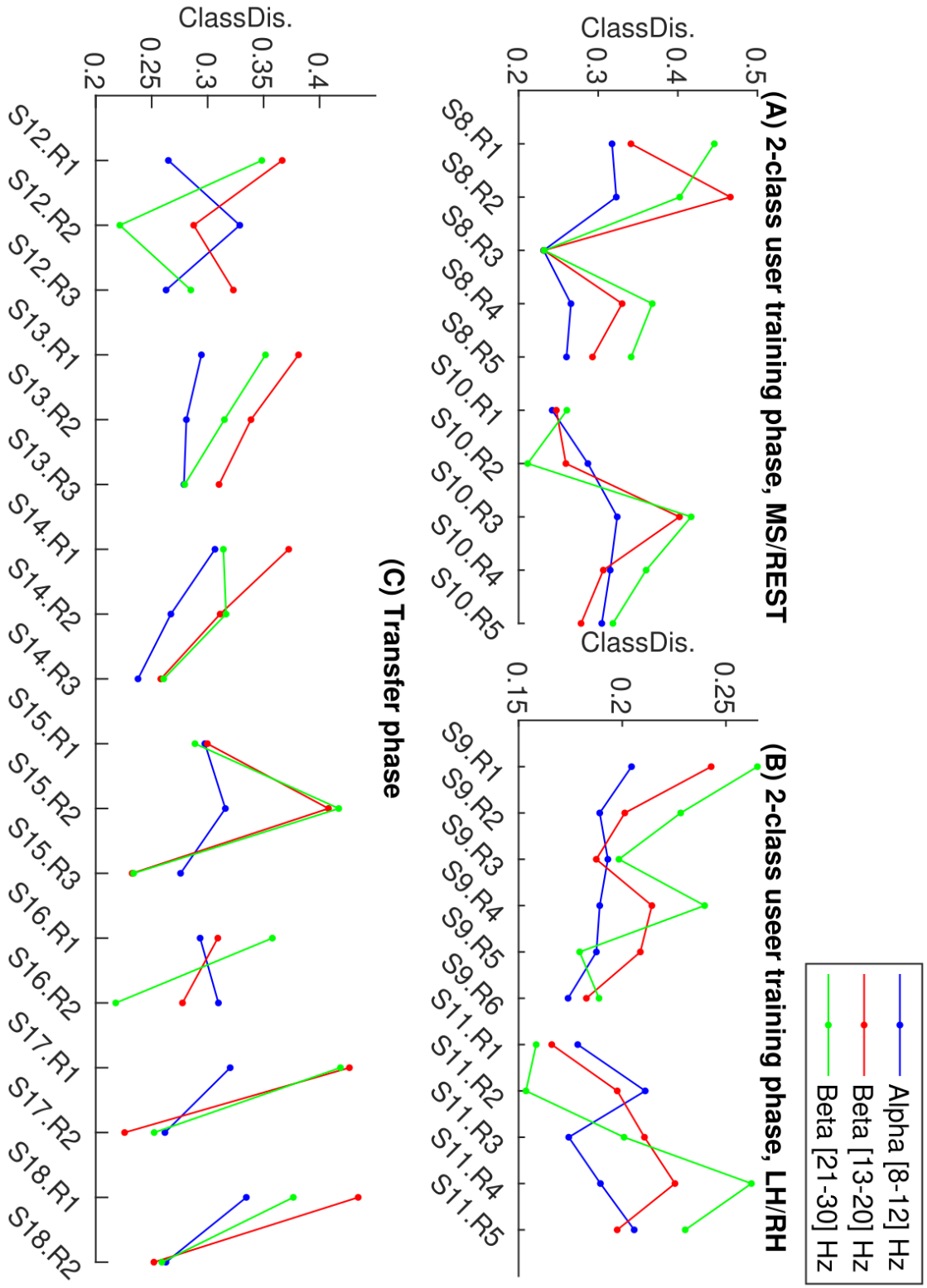


Figure 3.8: Class discrimination for each run of each session of **(A)** MENTAL-SUBTRACTION vs REST of two-class training phase, **(B)** LEFT-HAND vs RIGHT-HAND of two-class training, and **(C)** Transfer phase. Different colours represent different frequency bands, each point represent a run.

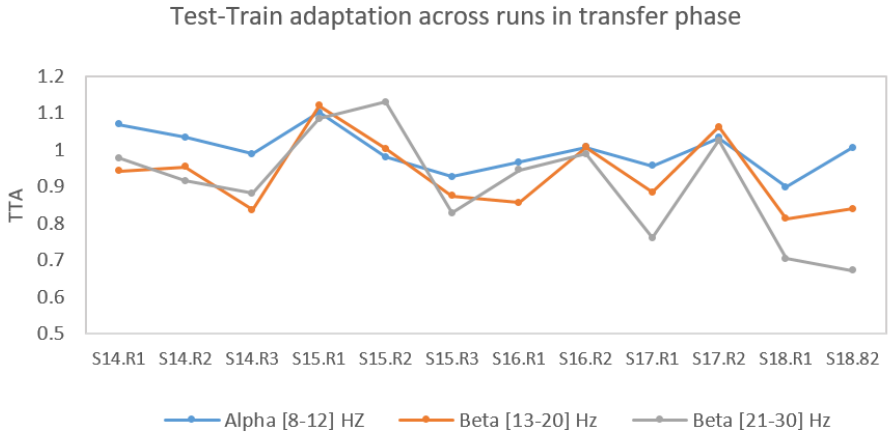


Figure 3.9: Evolution of the proposed test-train adaptation (TTA) metric over the runs in the transfer phase in different frequency bands.

new metric *TTA* in the different frequency bands across the different runs from session S14-S18, see Figure 3.9.

Here the (offline) training test was the data from Sessions S8-S13, and the online test set Sessions S14-S18. Indeed, the FgMDM classifier weights were optimized on this training set and kept fixed for this online test set. Pearson correlation between *TTA* and run index confirmed a significant negative correlation in the high β-band ($\rho = -0.57$, $p = 0.05$) with a mean TTA value of 1.00 for the first two sessions and of 0.79 for the last two sessions. In α and low β-band the correlations of TTA and run index were not significant though ($\rho = -0.44$, $p = 0.14$);($\rho = -0.26$, $p = 0.40$). Overall, this suggested that our pilot learned to increase his BCI classification accuracy not by producing increasingly more distinct EEG patterns (as measured by *classDis*), but by learning to produce EEG patterns that increasingly better matched those expected by the classifier for each class (as measured by TTA).

Game performance over the runs of each session, in terms of the required time to complete the game (in seconds), is illustrated in Figure 3.10.

For at least 4 of the 7 game sessions, we observed improvement in performance across runs, within-sessions. However, between-sessions, there does not seem to be any learning effect with improvement across sessions. Some sessions/runs were at times much better (notably session 17), but such performance was not sustained in subsequent sessions. This lack of apparent learning to control the game better was confirmed by the lack of significant correlation between game completion time and run index ($\rho = 0.17$, $p = 0.41$).

3.3.2 Competition results

Six international BCI teams participated in the 2019 CYBATHLON BCI competition in Graz during which the race completion time was the winning criterion. The maximum duration of the race was four minutes. If two (or more) pilots could not finish the race within the allotted time, the distance traveled was used to arbitrate the pilots. Victory was played in three rounds. Two qualifications rounds to select the three best pilots and a third round to determine the final ranking. Our pilot did not complete the races of the first two rounds in less than four minutes and therefore was not qualified for the track A finale (i.e., the finale with the best pilots) but the track B finale. During this finale, he did not finish the race in the allotted time (distance travelled: 399.8/500) and was ranked 5th out of 6. It should be mentioned that we faced technical issues during the competition. Firstly, instead of only two, the pilots had to go through four qualification races due to communication problems between the racing team and the organization's material which affected several pilots in the first

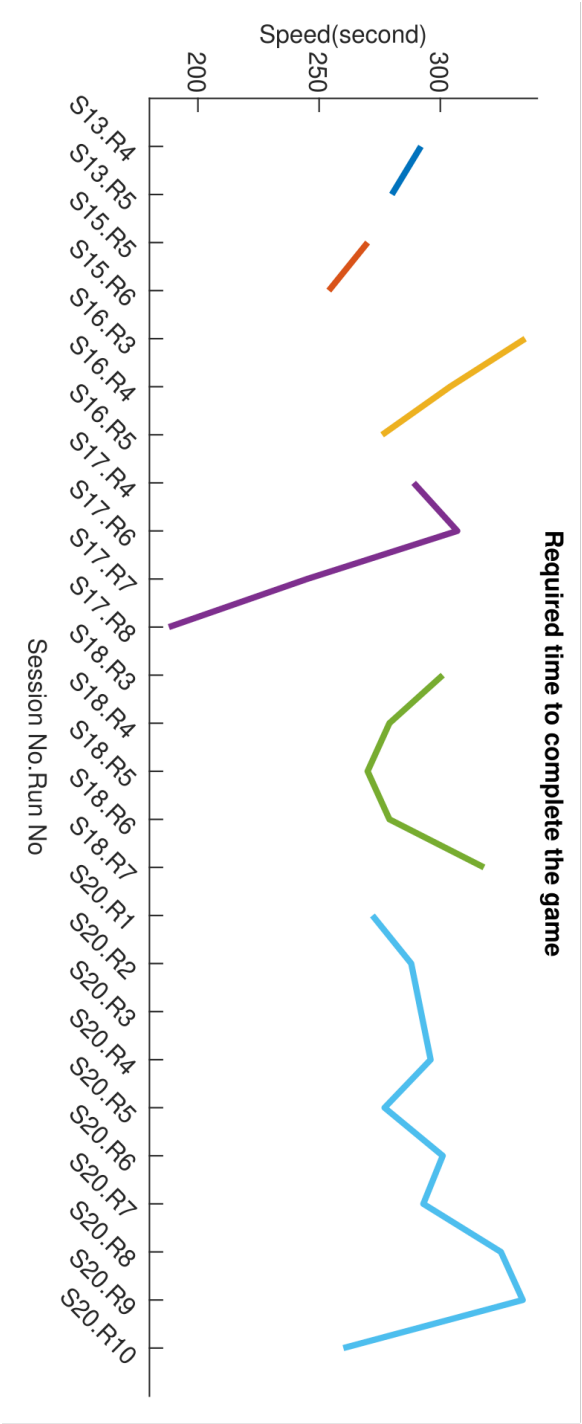


Figure 3.10: Game performance in terms of required time for completing the race. Different colours represent different sessions. Session 19 game times were not recorded.

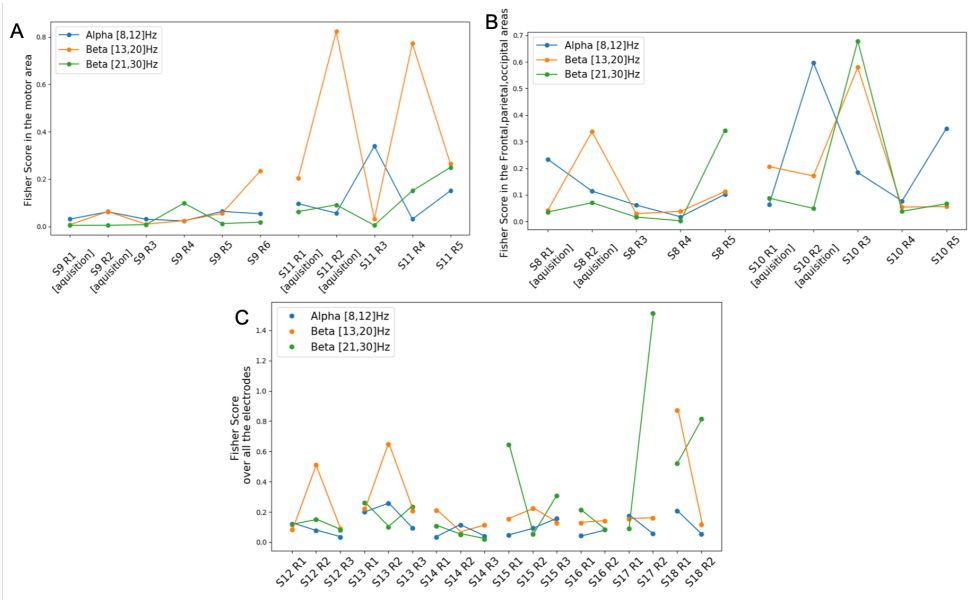


Figure 3.11: Mean Discriminability score (Fisher Score) for each frequency band for different conditions across the sessions. In A: Mean FS in the motor area (2-class motor imagery BCI) for three different frequency bands. In B: Mean FS in the frontal, parietal and occipital areas (2-class mental task BCI, i.e., REST vs MENTAL SUBTRACTION) for three different frequency bands. In C: Mean FS in all the electrodes (4-class BCI), for the three different frequency bands.

two races. Secondly, we had to change an EEG electrode on stage, between two races, as it was not working anymore.

3.3.3 Neurophysiological results

3.3.3.1 Exploratory phase

As mentioned before, the exploratory phase was more of an adaptation phase for both the pilot and the experimenters rather than a training phase. We used those 7 sessions to adapt the cap, the algorithms and for the pilot to be more familiar with the system. Thus, we did not do any neurophysiological analysis on that phase.

3.3.3.2 2-class training phase

During the 2-class training phase, we trained our pilot for four sessions in two weeks (i.e., sessions twice a week). Two sessions were dedicated to motor imagery BCI (LEFT- vs RIGHT-HAND movement imagination) and two to non-motor mental-task BCI (MENTAL SUBTRACTION vs REST). During the 2-class motor imagery BCI (Figure 3.11.A), we observed a learning effect in the low β-band (13-20 Hz in orange). Indeed, there is a significant correlation between run indexes and the FS in this frequency band (ρ=0.70, $p < 0.05$). No other significant correlation were observed.

3.3.3.3 Transfer phase

During the transfer phase in which we alternated between the game and a more classic 4-class BCI during 7 sessions, We can observe in Figure 3.11.C that the mean discriminability over all 46 electrodes seems to be increasing over time in both low and high β-bands (13-20Hz and 21-30Hz). However we did not observe any significant correlation between FS and the run number. Such results seem in line with the results obtained with *classDis* and *TTA* previously: our pilot did not seem to increase the discriminability of his EEG patterns with learning (as measured with FS or *classDis*), but rather to

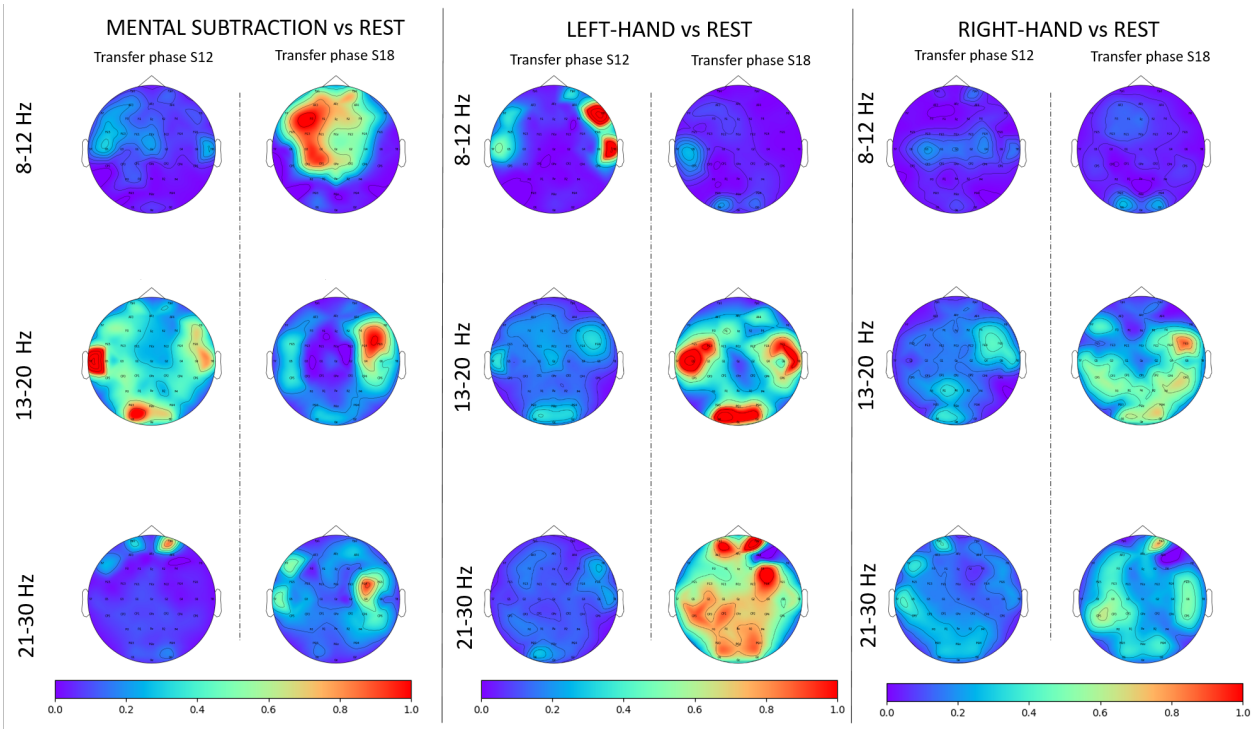


Figure 3.12: Topographic maps (Fisher Score) for each frequency band for the 3 active classes against rest state. For each condition, we compared the mean FS over the runs of the first session of the transfer phase and the one of the last session of the same phase.

produce EEG signals increasingly more systematically similar those used to train the classifier (as reflected by *TTA*).

In order to verify the soundness and relevance of the subject's class-wise EEG patterns, we used topographic maps of the Fisher Score for the 3 active classes against the rest class, during the transfer phase. We did so for each frequency band and compared the first session (i.e., mean of the FS over all runs) of this phase with the last one (Figure 3.12). Results showed that there was a learning effect in the expected brain areas from a neurophysiological point of view: FS increased in the frontal area for MENTAL SUBTRACTION vs REST. We also observed that, when comparing motor tasks with REST, there was an increase in FS in the motor areas (around C3 and C4), as well as in visual areas. However, we did not observe any lateralized pattern, but rather a bilateral one. This would explain why our pilot had difficulties to produce distinct EEG patterns for the two hand motor imagery tasks.

3.3.4 User experience

3.3.4.1 Questionnaires

The questionnaires described in Section 3.2.3.2 provide us with scores for 5 factors, i.e. mood, motivation, mindfulness, agency and cognitive load (the last two being assessed post session only). For the sessions in which the questionnaires were administered, the results are reported in Figure 3.13.

For the pilot, the first session of the transfer phase (12) was particularly difficult. He encountered a rather low online classification performance and reported significant fatigue. As shown in Figure 3.13, this resulted in a particularly low mindfulness at the end of this session (12), as well as a rather low mood both at the end of this session and at the beginning of the next one (13). Regarding motivation, although quite variable from one session to the next, remained overall rather high throughout training as reflected by the interview. Also, a high level of agency seems to be associated with high peak times when playing the game (cf. Figure 3.10).

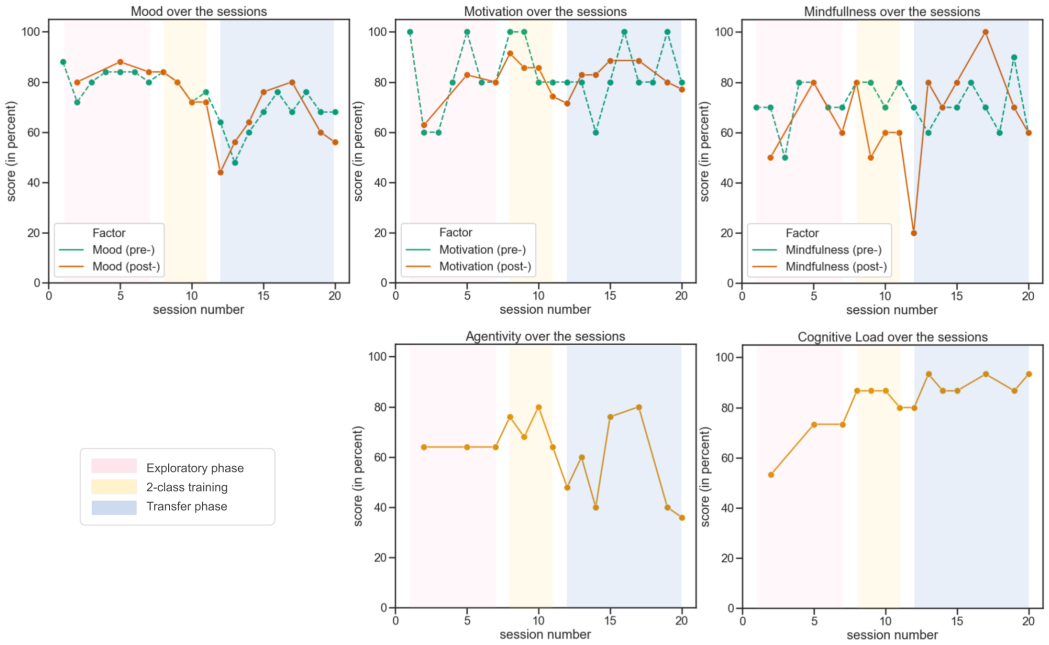


Figure 3.13: Evolution over sessions of user experience metrics, i.e. user mood, motivation, mindfulness, agency and cognitive load.

Finally, during the course of the training, the cognitive load gradually increased and was at its highest during the transfer phase (4-class user training and game).

3.3.4.2 Interview

A semi-structured interview of the pilot was conducted after the competition in order to assess different dimensions of user experience including acceptability, acceptance and satisfaction. The pilot's answers provided herein-after are organised based on the four main steps of the "adventure", i.e., i) before the BCI training, ii) during the BCI training, iii) during the competition and iv) after the competition.

During the interview, the pilot indicated that he had never used a BCI before. When he was offered to take part in the CYBATHLON, he first felt curious and considered this opportunity as challenging and "awesome". He was extremely attracted by the idea of integrating a research project, to contribute to the emergence of a technology, from the very beginning of the process. Being part of this new "adventure", culminating with the CYBATHLON competition, was his main motivation. He also mentioned that it was very easy for him to envision many promising applications of BCIs for the general public. However, he did not mention having any expectation related to the improvement of his clinical condition. Rather, he was excited to be part of a team, represent his country in an international competition, meet with new people and discover a field of which he knew nothing before. The pilot's close family and friends also strongly supported him in his project. Thus, the pilot's BCI acceptability a priori was high. He was very motivated, with a high perceived usefulness of the technology and very positive representations. Moreover, He will mention later in the interview that he did not expect the training to be so hard, showing that he also perceived BCIs as easy to use. This high acceptability resulted in a great engagement of the pilot in the training procedure, which is reflected by the results of the questionnaires depicted in Section 3.3.4.1.

During the second phase of the interview, dedicated to the BCI training phase, the pilot mentioned that keeping in mind the final objective, namely the competition in Graz, enabled him to keep his motivation high all along the training program. Nonetheless, it was not easy as he had under-estimated the mental effort required for completing BCI sessions. This resulted in huge tiredness after some sessions. He also found it difficult to manage, from a psychological standpoint, the between-session variability in terms of performances as he felt like he had no control over it. This second difficulty led

to high frustration levels during and after some of the sessions that he has had to learn to master. The pilot mentioned that beyond the objective of the competition and the trip to Austria with his friends, the emotional as well as the cognitive support provided by the team members, who were much present and explained the potential origins of variability in performance, helped him to remain motivated.

As the CYBATHLON competition approached, the pilot reported feeling both extremely excited, impatient and a bit stressed. He also felt the stress of the team, especially when technical issues occurred on stage. During the race, the pilot switched between frustration and relief periods. But in the end he was mainly proud and happy. Most importantly, he insisted on the fact that he loved the way the competition was organized. It was a real show, with an over-excited audience and interviews. He felt like “any other athlete, on equal terms". He adds, “I was not a disabled person helped by researchers. I was the pilot of a team". the pilot qualified this feeling as “refreshing" and wished that this kind of event where disability is not associated with pathos and sentimentality were more frequent.

Now that the “Adventure" is over, the main feelings of the pilot when he recalls the whole process are a huge satisfaction and happiness. It was a bit hard at the beginning not to have the training routine anymore. Now, he would like to compete again and increase his performance. He thinks that mental coaching and meditation sessions could help. Beyond the CYBATHLON experience, the pilot affirmed that he would recommend non-invasive BCIs to anyone who is “ready to train hard". He also insisted on the importance to explain first how this (impressive) technology works in order to play down the assumed danger/difficulties associated with these technologies. He also mentioned that he would ready himself to train hard to learn to use them for controlling assistive devices in the future. He is certain that non-invasive BCIs will be made accessible and reliable for home automation notably in the near future. This shows that the pilot’s experience with BCIs resulted in a high acceptance, associated with a high perceived usefulness and a certain confidence in the fact that these technologies will soon be reliable and easy enough to be usable and useful outside of laboratories.

3.4 Discussion

During this study that lasted about three months, we conducted multiple BCI experiments with a tetraplegic user both inside and outside the laboratory.

This encouraged us to tackle multiple challenges associated with BCI use over multiple days, in “real life", with an actual end-user. Challenges were related to non-stationarity problems, user training but also managing the short time we had before the competition.

At the beginning of the transfer phase (Sessions 12 and 13), even when using a classifier trained on multiple runs and sessions, the resulting classification accuracy still suffered from a high variability between runs and sessions. Using a newly proposed adaptive approach to model both within and between-session variabilities, from Session 14 onwards, led to improvement in classification accuracy across runs and sessions. This observation confirmed the effectiveness of the proposed adaptive Riemannian classification approaches for addressing non-stationarity effects.

During the training, we observed user learning across sessions, notably in the transfer phase, and particularly when using the adaptive approach with the bar feedback. However, we did not observe any clear increase in game performance across sessions for controlling the CYBATHLON BCI game. We only observed within-session game performance improvement, but such improvement did not sustain to the subsequent sessions. Within-session game improvement was probably due to the within-session adaption of our classifier, which was progressively adapting the classifier to the game context. Indeed the classifier was trained on the bar-feedback context, and not the game context. Indeed, the asynchronous nature of the game prevented us from using it to collected ground truth EEG data to calibrate the BCI. However, the differences between the standard bar-BCI training and the CYBATHLON game training, e.g., the continuous vs discrete feedback, the positive-only vs positive and negative feedback or the simple vs complex visuals would likely lead to change in EEG patterns. Overall, this result thus confirmed results from [88] which stated that both standard feedback BCI

training and application game training were different tasks, that both needed a dedicated training, possibly both for the user and for the machine (classifier).

In order to avoid impeding the user learning, it thus seems necessary to provide an accurate and stable feedback. In our study, unsupervised adaptive methods - that model global EEG changes within and between sessions, seemed useful to improve feedback stability and BCI performances.

This proposed adaptive approach only models and corrects global EEG distribution shifts within (using adaptive rebias FgMDM) and between sessions (using between-session baseline projection) but not task-related EEG changes. This way, we hope it only contributed to stabilizing BCI feedback, but not in making the user "lazy", by enabling him to overly rely on the ongoing machine learning doing all the work, as observed in [120]. Therefore, using the proposed adaptive approaches, which reduce the shift within and between sessions, we probably reduced the risk of confusing the user with a continually changing feedback. Interestingly enough, the statistically significant improvement that we observed in both online classification accuracy and user adaptation to the BCI classifier training set (TTA metric), together with the non-significant improvement in *classDis* provide some evidence for a relatively new type of user learning. Indeed, our user learned to adapt his EEG signals to the BCI classifier, instead of increasing his EEG separability between classes (the typically studied type of BCI user learning). This potentially open doors to new ways to assess and study BCI user learning in the future. Naturally, this result is here shown for a single subject, and would thus need to be further studied with several other subjects, and other BCI designs, e.g., with non-adaptive classifiers.

As shown in Figure 3.13, the pilot's cognitive load showed an increasing trend throughout the training. At first, this may seem surprising since task execution supposedly requires less load as expertise increases. Throughout progressive training, one might thus have expected a more stable or even decreasing load: the training could build and automatize patterns associated with MTs [123] thereby freeing up processing resources for the execution and learning of new aspects of training exercises, e.g., for controlling the game. We can first hypothesize that the decrease in the load relative to increased expertise was lower than the increase associated with the introduction of more complex and demanding exercises. Another line of interpretation would be to consider the causal factors of the cognitive load related to the environment. For example, stress, emotions and uncertainty can limit working memory capabilities [124] and thus increase cognitive load and impede learning. An acceleration of the training schedule (more sessions per week as the competition was growing nearer) and the approach of the competition might have triggered this type of process. It might therefore be recommended to carefully assess these elements in the future and to develop methods to support the user and help them to reduce the high load resulting from stress, emotions or uncertainty.

3.5 Lessons learned

The CYBATHLON BCI series 2019 was, for our team, a wonderful opportunity to experiment a long term BCI training with a end-user and a clear final objective and deadline: the CYBATHLON BCI competition. Yet, it was a first for our team and we learned a lot from our mistakes.

First, due to the short schedule, decisions had to be taken quickly without being able to step back. We believe that we took final decisions quite fast that were not the most relevant ones. One of those decisions was the tasks the pilot had to perform to control the BCI and more specifically the motor imagery tasks. Indeed, as our pilot was tetraplegic since 2007, it was not easy for him to do those tasks but also to generate distinct brain signals for those two tasks (i.e., LEFT- and RIGHT-HAND movement imagination). Spending more time on screening could have help us choosing tasks he was more comfortable with but also that were easier to discriminate.

Another challenge we had to face was the design and implementation of new algorithms while training our pilot such as using Riemannian geometry online and addressing the issue of non-stationarity effects. It was, at times, confusing and stressful for both the pilot and experimenters to train and test at the same time.

Regarding our adaptive approaches, for the game, the number of trials for the different classes was not equal, with the NOINPUT class (rest) being over represented (once a command was issued, the

pilot could rest). This could have biased our within-session adaptive approach towards that class. We need to address this issue in the future.

Moreover, the pilot's testimony regarding the support of the team highlights three very important aspects that should not be neglected. Indeed, it seems necessary that researchers/clinicians clearly introduce and discuss with the user/patient in advance the system, the way it works, the level of performances that can be expected, the duration and potential difficulty of the training, etc. It is also of utmost importance to provide a comforting presence all along the training phase. Finally, defining clear objectives will most likely favour the engagement of the user/patient, provided that the two previous recommendations of cognitive and emotional support are followed.

Overall, the whole training protocol included 20 training sessions, arranged in three phases. This was mostly due to its exploratory nature. If we were to do it again, we would make the exploratory phase much shorter than the 7 sessions performed here. We would focus on screening various MTs in a systematic way to identify faster and better the ones to further train in the subsequent phase. The 2-class training phase appeared as useful to train our pilot in a progressive way. Indeed, an initial test with all 4-classes at once quickly overwhelmed our pilot. He then appreciated training first with pairs of MTs only. Finally, for the transfer phase, ideally, we should make that phase longer - to train the pilot more, and include game training earlier in that phase.

Finally, the environmental conditions during the competition were far from the ones we had at the pilot's home or in the lab. Indeed, during the training we were making sure that there was no disturbing element such as noise or movement in front of the pilot. In contrast, during the competition, the public was supporting pilots and therefore moving and making noise. In addition to the stress due to the competition, we were far from the training conditions. In the future, training for the CYBATHLON competition should also include stress management and BCI training in noisy and stressful conditions.

All the points above certainly contributed to our results at the actual CYBATHLON competition, where we ended up 5th out of 6. Interestingly enough, except team NITRO 2 (our team was team NITRO 1), which ranked 6th/6, all the other teams already had at least one previous experience in this competition (for the CYBATHLON 2016 and/or the CYBATHLON BCI series 2015) with their pilot, who thus trained to control a BCI for at least one year (our pilot trained for 3 months), sometimes much longer. This seems to also suggest that to achieve good BCI design and training control, both BCI scientists teams and BCI CYBATHLON pilots need substantial training and experience, which we lacked for our first CYBATHLON attempt.

3.6 Conclusion

Our results showed a learning at all levels (i.e., user, machine, and experimenters). Indeed, during the few months of training we were able to observe a user learning. Indeed, the classification accuracy significantly increased during training, while our proposed metric, i.e., test-train adaptation (TTA) significantly decreased, both reflecting user learning. In addition, we were able to propose a new Riemannian adaptive approach to reduce EEG distribution shifts within and between sessions. Such an approach could, in the future, be used to improve the user learning by stabilizing the BCI feedback. Controlling a BCI can be long and difficult as generating distinct brain signal that the BCI can recognize is a skill to be learned and we believe that improving the technology to help the user achieve that goal is one of the research area we need to focus on. In addition to the improvement of the technology, our study showed that focus on user training and user experience is also essential. Finally, we also learned a lot by doing mistakes during the training. This enabled us to identify several interesting research directions for the future.

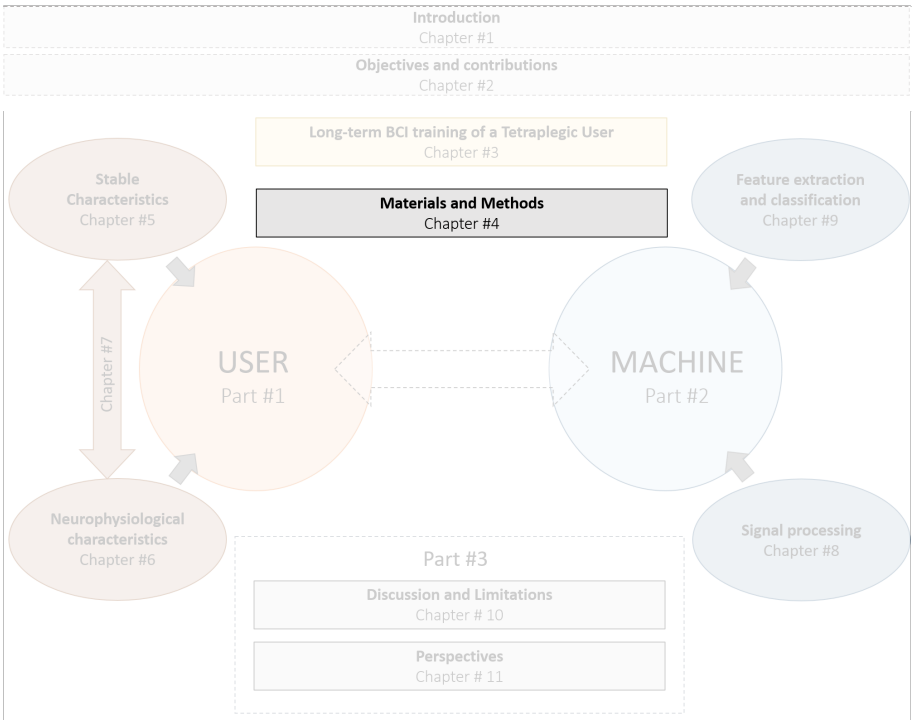
This experiment revealed the importance of being able to predict MT-BCI performances of a user to implement the most optimal subject-specific training protocol. Indeed, during this study, we moved forward blindly, making assumptions and modifying the training quite often. A first phase dedicated to 1) the extraction of features used in MT-BCI models of performances and 2) the prediction of future MT-BCI performances using those models could have helped us designing a relevant training protocol.

This process could have affected several parameters: the motivation of both users and members of the reseach team, performances and the general organisation of the training phase. The following Parts of this manuscript are therefore dedicated to the understanding of MT-BCI performances and the identification of factors influencing it using predictive models of performances across experiments and datasets.

4

Materials and methods

ROADMAP -



QUICK SUMMARY -

The following chapter will provide a detailed description of the 5 datasets (2A, 2B, 3A, 3B, 3C) used in our different studies. All of them are derived from experiments using MT-BCI that were designed to study user-training. In this PhD manuscript, those datasets were used to 1) extract predictors, related to the user (Part I) and the machine (Part II), of BCI performances and 2) build model of prediction using those predictors. This chapter will provide details regarding materials and methods used for the data acquisition :

- the participants (number, age)
- the characteristics of the experimental protocol (number of trials, runs, session; feedback provided; Mental tasks)
- the EEG recordings and pre-processing

In addition we will give a detailed description of the machine learning method used to build models of prediction, and more precisely, the Least absolute shrinkage and selection operator (LASSO). Computational modeling is used to simulate and study complex systems using mathematics, physics and computer science. In this thesis, we want to simulate MT-BCI performances to better understand user learning.

Summary tables (one for data acquisition and one for machine learning methods) will briefly recapitulate these elements in each chapter. The reader can then refer back to this chapter for detailed explanation of each element presented in the table.

Contents

4.1 Introduction

This section introduces the materials and methods used: 1) for the data acquisition in the different studies led in the context of this thesis 2) for the creation of models of prediction. The previous Chapter highlighted the importance to better understand user learning to provide the user the most suitable training. For this purpose, Parts I and II will be theoretical and computational studies using different datasets. Therefore, as all our datasets were designed to study user-learning when using MT-BCI and had a large number of similarities we decided to describe them in a same chapter. In total, we used 5 different datasets, two (datasets 2A and 2B) were based on experiments using MI-BCI with two MI tasks: Left-hand motor imagery (LH) and Right-hand motor imagery (RH) and three (datasets 3A, 3B and 3C) were based on experiments using MT-BCI with three MTs: LH, mental rotation (ROTATION) and mental subtraction (SUBTRACTION). The initial purpose of each dataset experiment will be described in this Chapter. In addition, we will describe all the tests applied for user profile evaluation that will be used in Chapter 5, focusing on the identification of predictive factors of MT-BCI performances derived from users' traits (i.e., stable characteristic). Datasets 3A, 3B and 3C will be used in those analyses. Then, we will describe EEG recordings and pre-processing methods used in all datasets. Data-driven methods used in those protocols will be further studied (but also described) in Part II.

Datasets 2A and 2B, using a 2-class MI-BCI, will be used in Chapters 6, 7, 8 and 9.

Finally, as we used the same machine learning method (i.e., LASSO) [125] to build computational models, a section will be dedicated to the description of this method. We chose the LASSO as it promotes sparse solutions and thus allows us to have interpretable models.

One may note that while the general principle of all the experiments was the same (i.e., it was based on the Graz training protocol, MT-tasks), each experiment had some specificities in terms of training protocol and signal processing. Thus, in the next Chapters brief tables will present the materials and methods used for the data acquisition but also the computational modeling.

4.2 User profile evaluations

The user profile evaluations aimed at determining potential relationships between the users' traits (i.e. stable characteristics) and their BCI performance (see Chapter 5). Only the questionnaires used in our research will be presented in this section.

4.2.1 Personality profile

In order to assess different aspects of the personality of the users we used the 16 Personality Factors 5 (16PF5) [126]. It is a self-report personality test that provides a score for sixteen primary factors of personality (warmth, reasoning, emotional stability, dominance, liveliness, rule consciousness, social boldness, sensitivity, vigilance, abstractness, privateness, apprehension, openness to change, self-reliance, perfectionism and tension) as well as five global factors of personality (extraversion, anxiety, tough mindedness, independence and self control). See Appendix Afor details about those factors. The test is composed of 185 forced-choice questions in which the participant must choose one of three different alternatives. The test was done online and lasted between 30 and 45 min.

4.2.2 Cognitive profile assessment

We used a mental rotation test measuring spatial visualisation abilities (SVA). We used the french version [127] of the mental rotation test by Vandenberg and Kuse [128] that was used in previous BCI studies [52]. This is a mental rotation test inspired by the work of Shepard and Metzler [129]. The test (paper and pencil) is presented in the form of 20 items divided into two sets of 10 items. Each set has to be completed in three minutes maximum. An item consists in a 3D shape on the left and four 3D shapes on the right. Among the four 3D shapes, only two are similar to the left one with

a rotation around the vertical axis. For each item, participants are asked to select the two correct answers out of the four possibilities.

4.2.3 Others

Among the online questionnaires participants had to complete during BCI sessions one retrieved demographic & biosocial data (age, gender, education level, etc.).

4.3 Participants

4.3.1 2-class MI-BCI datasets

Dataset 2A and 2B use data from two different experiments [130, 131]. They were all based on the same BCI paradigm, i.e., MI-BCI with two motor imagery tasks: LH and RH. All participants were BCI naïve and none of them received a financial compensation.

The purpose of dataset 2A experiment [130] was to evaluate if the experimenters' gender influenced MI-BCI user training outcomes, i.e., performances and user-experience. The experiment involved 6 experimenters (3 women and 3 men) each training 5 women and 5 men (60 participants in total) to perform right versus left hand MI-BCI tasks over one session. The experiment of dataset 2B [131] is a contribution to this PhD work and was designed to evaluate how adding constraints on a data-driven algorithm used during the system calibration could impact the performances of BCI users. Twenty-one (21) participants were included in this dataset. More details about the implementation of the corresponding experiment will be given in Chapter 8. The 2 studies were conducted in accordance with the relevant guidelines for ethical research according to the Declaration of Helsinki and approved by the ethical committee of Inria (the COERLE, Approval number: 2018-13).

4.3.2 3-class MT-BCI datasets

Dataset 3A, 3B and 3C use data from three different experiments [52, 64, 132, 133]. They were all based on the same BCI paradigm, i.e., MT-BCI with three mental tasks: LH, ROTATION and SUBTRACTION. All participants were BCI naïve and only participants from dataset 3A, 3B and 3C received a financial compensation.

The purpose of dataset 3A experiment [52], was to determine how the users' cognitive and personality profiles influenced their MT-BCI performances. For this experiment, 18 subjects (9 women, 9 men; aged 21.5±1.2 year-old) took part in 6 MT-BCI sessions, on 6 different days with a mean number of 4±2 days between two sessions. The experiment of dataset 3B [64, 132] was designed to assess the influence that a Spatial Ability (SA) training procedure might have on MT-BCI performances. Fourteen subjects (8 women, 6 men; aged 22.6±4.6 year-old) took part in this study. Each of them did 3 MT-BCI training sessions and 3 other cognitive training sessions (without BCIs), which consisted either in an SA training (7 subjects) or in a verbal comprehension training (7 subjects) procedure. There was, on average, 18.8±4 days between two BCI sessions. This relatively long time was due to the fact that the cognitive training without BCI was conducted between the first two BCI sessions. Finally, in dataset 3C [133], 10 subjects (5 women, 5 men; aged 20.7±2.1 year-old) were accompanied by a learning companion called PEANUT (personalized emotional agent for neurotechnology user-training) providing social presence and emotional support during 3 MT-BCI training sessions. The goal of this experiment was to evaluate the influence that a social presence and emotional support had on had on MT-BCI performances. The 3 studies were conducted in accordance with the relevant guidelines for ethical research according to the Declaration of Helsinki and approved by the ethical committee of Inria (the COERLE, approval numbers: 2015-004, 2016/02-00, 2016/02-00).

4.4 Experimental Protocol

First, in experiments 2A and 2B only, EEG baselines (resting state) were recorded with eyes-open and eyes-closed condition (2 3-minute baselines). The baselines were followed by the BCI session itself. Then BCI the training was composed of

- one 2 hours MI-BCI session of six 7-minutes BCI runs with N_{trials}=40 trials per run (20 per task) for dataset 2A and 2B.
- six 2 hours MT-BCI sessions (on six different days) of five 7-minutes BCI runs with N_{trials}45 trials per run (15 per task) for dataset 3A.
- three 2 hours MT-BCI sessions (on three different days) of five 7-minutes BCI runs with N_{trials}45 trials per run (15 per task) for dataset 3B and 3C.

during which participants have to learn to control a BCI. The learning protocol used for this experiment was the Standard Graz BCI Training Protocol [103] or an adaptation of it.

4.4.1 The Standard Graz BCI Training Protocol

During each BCI run, participants had to perform N_{trials} presented in a random order, each trial lasting 8s (see Figure 4.1). At t = 0s, a cross was displayed on the screen. At t = 2s, an acoustic signal announces the next appearance of a red arrow, which then appears one second later (at t = 3s) and remains displayed for 1.250s. The arrow points in the direction of the task to be performed. Finally, at t = 4.250s, a visual feedback was provided in the shape of a blue bar, the length of which varied according to the classifier output. Only positive feedback was displayed, i.e., the feedback was provided only when there was a match between the instruction and the recognised task. The feedback lasted 3.75 s and was updated at 16Hz, using a 1s sliding window. After 8 seconds of testing, the screen will turn black again. The participant then rests for a few seconds, and a new cross is then displayed on the screen, marking the beginning of the next test.

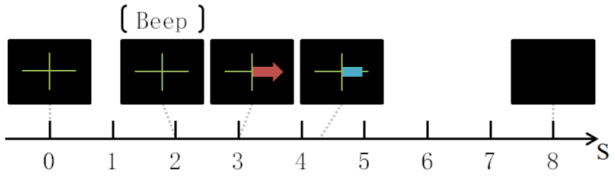


Figure 4.1: Timing of one trial in the Graz Protocol

In dataset 2A and 2B, the first two runs were used as calibration in order to provide examples of EEG patterns associated with each of the MI tasks to the system whereas in dataset 3A, 3B, 3C the calibration was on the first run of the first session and the bias of the shrinkage Linear Discriminant Analysis (sLDA) was recentered at the beginning of each session, using the EEG data from the first run of that session (see Section 4.5 for more details about the pre-processing pipline). As the classifier was not yet trained to recognise the mental tasks being performed by the user during the calibration run(s), it could not provide a consistent feedback. In order to limit biases with the other runs, e.g., EEG changes due to different visual processing between runs, the user was provided with an equivalent sham feedback, i.e., a blue bar randomly appearing and varying in length, and not updated according to the classifier output.

4.4.2 Tasks and instructions

Left-hand motor imagery (LH) and Right-hand motor imagery (RH) tasks refer to the kinesthetic continuous imagination of a left- or right-hand movement(respectively), chosen by the participant, without any actual movement [134]. ROTATION and SUBTRACTION tasks correspond respectively to the mental visualisation of a 3 Dimensional shape (of their choice) rotating in a 3 Dimensional space [134] and to successive subtractions of a 3-digit number by a 2-digit number (ranging between 11 and 19), both being randomly generated and displayed on a screen. For the two class studies (2A and 2B) participants had to control the BCI using the two motor-imagery (MI) tasks. In the 3-class studies, participants were asked to learn to perform 3 mental tasks (MT), namely LH task, mental rotation (ROTATION) and mental subtraction (SUBTRACTION), which were chosen according to Friedrich et. al, [134] who showed that these tasks were associated with the best performance on average across BCI-users. Participant were allowed to chose either the same imagined movement for the LH and RH tasks or two different ones and were free in the choice of their movement [135]. Finally, participants were encourage to perform a kinesthetic imagination during MI tasks [136] (i.e., that requires the feeling of sensations such as pressure, acoustic and olfactory stimuli, and muscle tension). Finally, participants were instructed to try to find the right strategy so that the system shows the longest possible feedback bar in the correct direction (i.e. left or right).

4.5 EEG recordings and pre-processing

For the 2-class datasets (namely 2A and 2B), EEG signals were recorded using 27 active scalp electrodes (Fz, FCz, Cz, CPz, Pz, F3, FC5, FC3, FC1, F4, FC2, FC4, FC6, C5, C3, C1, C6, C4, C2, CP5, CP3, CP1, CP6, CP4, CP2, P3, P4) 10-20 system, active electrodes, g.USBamp amplifier - g.tec, Graz, Austria) and processed online using OpenViBE [**?**]. To those electrodes we added two EMG (ElectroMyoGraphy) electrodes (left and right hand wrist) and 3 EOG (ElectroOculoGraphy) electrodes to record artifacts (below, above and on the side of an eye).

For the three 3-class MT-BCIs datasets (3A, 3B and 3C), EEG signals were recorded using the same system as above with 30 active scalp electrodes (F3, Fz, F4, FT7, FC5, FC3, FCz, FC4, FC6, FT8, Oz, C5, C3, C1, Cz, C2, C4, C6, CP3, CPz, CP4, P5, P3, P1, Pz, P2, P4, P6, PO7, PO8) .

All data was recorded, processed and visually inspected with OpenViBE [**?**]. OpenViBE is a free and open source software enabling to design a BCI without programming: through the use of pre-existing signal processing and machine learning functions that can be connected using graphical programming.

We used data from two calibration runs to build users' specific classifier in the two-class datasets (2A and 2B) and one calibration run in the three-class datasets. The EEG signal processing pipeline used to classify the two motor imagery tasks online was the same for the two datasets 2A, 2B with the exception of the use of a different optimization algorithm to select the most discriminant frequency band (MDFB) before the spatial filtering of the signals using 3 pairs of Common Spatial Pattern (CSP) filters [137], to discriminate one class versus the other. In dataset 2A, we used the algorithm of [138] (Algorithm 1 in the corresponding paper). This algorithm selects the frequency band whose power in the sensorimotor channels that maximally correlates with the class labels. In dataset 2B, we modified this algorithm by adding constraints to select an MDFB width larger than 3.5Hz and an MDFB mean value under 16Hz (see Chapter 8 for more details about this contribution). For the three-class datasets (3A, 3B, 3C) the EEG signal processing pipeline used to classify the three mental tasks online was the same. EEG signals were first band-pass filtered in 8-30 Hz, using a 4^{th} order Butterworth filter. Then they were spatially filtered using 3 sets of 2 pairs of Common Spatial Pattern (CSP) filters, each set being used to discriminate one class versus the other two.

Finally in all datasets, the log band power of the spatially filtered EEG signals were classified using a sLDA classifier [139]. The bias of the sLDA was recentered at the beginning of each session in the three-class datasets, using the EEG data from the first run of that session.

4.6 Computational models using Least absolute shrinkage and selection operator

One of the principal objective of our work was to create computational models that could predict and/or explain MT-BCI performances to better understand user-learning. Therefore, the variable to predict was MT-BCI performances predictors included both intrinsic user-related factors (i.e., user's traits and neurophysiological characteristics, Part I) and extrinsic factors (i.e., characteristics extracted from machine learning algorithms, Part II). Furthermore, in order to understand our results, models had to be interpretable. Finally, to identify relevant factors (among all the available ones) that could impact MT-BCI performances, our model had to select a reduced subset of variables. Therefore, for all these reasons, we decided to use a LASSO regression [125].

The LASSO regression is a method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the resulting statistical model. Indeed, it uses an L1 norm regularization with a penalty parameter λ (see Equation 4.1) that promotes sparse solutions, i.e., that selects only a small number of variables (many coefficients will be zero using this regularization). It is particularly adapted to reduce the number of relevant features when those features are more numerous than the training data [140] and enables the creation of interpretable models. As for any linear regression set up, we have a continuous output vector $Y \in \mathbb{R}^n$, a matrix $X \in \mathbb{R}^{n \times p}$ of p normalized features using a z-score normalization for n examples (the subjects) and a coefficient vector $\beta \in \mathbb{R}^p$ (the regression weight). The LASSO estimator is defined as:

$$\beta_{lasso} = \text{argmin}_{\beta \in \mathbb{R}^P} ||Y - X\beta||_2^2 + \lambda ||\beta||_1 \tag{4.1}$$

where, $||u||_2^2 = \sum_{i=1}^{n} u^i$ for $u \in \mathbb{R}^n$ and $||\beta||_1 = \sum_{j=1}^{p} |\beta_j|$. For some values of the penalty parameter λ, some components of β_{lasso} will be set exactly to 0. Once β_{lasso} obtained, the variable to predict of the i^{th} user, V_{pred}^i, can be predicted from features extracted from the user x_i, as $V_{pred} = \beta_{lasso}^T x_i$ (T denotes transpose). In order to evaluate the stability and reliability of the different models, we used a leave-one-subject-out cross validation process. We also used an inner cross-validation (total number of training subjects: N-2) to automatically find the optimal λ among 100 possible $\lambda \in [0.001; 10]$, i.e., the one that minimizes the mean absolute inner-cross validation error and provides us with sparse features. We used this optimal λ to build a regression model and then, the outer cross-validation (N-1 subjects for training, 1 for testing) is used to evaluate this model.

For each cross-validation, different features were selected with different weights. Studying these features and weights first allowed us to assess the stability of the results. In a second step, we determined the reliability of the models by testing each of them on the participant not included in the training set during the cross-validation process. We then computed the mean absolute error of all the models, i.e, $\sum_{i=1}^{N} \frac{|V_{pred(i)} - V_{real(i)}|}{N}$, n being the total number of models generated.

In order to determine whether the models could make better-than-chance predictions, we estimated the empirical chance level in terms of mean absolute error, given our data. First, we randomly permuted the variable to predict of the training set, while keeping features identical, thus breaking the relationship between predictors and the variable to predict. Then, we used the LASSO as explained above to predict the variable of the left-out subject. We repeated this process 1000 times and stored each mean absolute error to obtain the distribution of the chance prediction. Then, we sorted those values in descending order and the 99^{th} and 95^{th} percentiles were used to identify the chance level for the mean absolute error for $p = .01$ and $p = .05$ respectively.

Note that this method presents some limitations. The LASSO does not necessarily lead to a unique model. Indeed, if several explanatory variables are highly correlated, the LASSO may randomly chose one for its model, which might make the resulting model potentially unstable in this case [125]. Another techniques i.e., Elastic Net (EN) [141] can lead to a unique model. Nonetheless, it does not enable a variable selection as sparse and therefore as interpretable as with the LASSO. We have also explored this method, as a sanity check, in order to find predictive models of performances. The EN method as well as results obtained with it are exposed in Appendix D.

Now that all the different elements of the materials and methods used in our studies have been introduced, the next Chapters will report the experiments themselves. For each of these experiments, the materials and methods will be reported using a table as the one that follows.

	DATASETS	2A*	2B**	3A***	3B***	3C***
GENERAL INFORMATIONS	Number of subjects	60	21	18	15	10
	Number of BCI sessions (Number of online runs per session)	1 (4)	1(4)	6(4)	3(4)	3(4)
	Number of Mental Tasks	2	2	3	3	3
	Mental Tasks[1]	LH-RH	LH-RH	MR-MC-LH	MR-MC-LH	MR-MC-LH
	Number of trials per tasks per run	20	20	15	15	15
	Additional EEG recordings	Two 3 min resting state baselines (open and closed eyes)	Two 3 min resting state baselines (open and closed eyes)	-	-	-
CLASSIFIER GENERATION	Number of calibration runs	2	2	1	1	1
	Signal processing parameters	Frequency band selection Algorithm	Frequency band selection Algorithm	Band pass filtering 8-30 Hz	Band pass filtering 8-30 Hz	Band pass filtering 8-30 Hz
	Features selection	CSP	CSP	CSP	CSP	CSP
	Classifier	LDA	LDA	LDA	LDA	LDA
PROFILE ASSESSMENT	Personality Profile	❖ 16 Personality Factors-5 ❖ Learning Style Inventory				
	Cognitive Profile	❖ Mental Rotation test				

[1] LH: Left Hand motor imagery; RH: Right Hand motor imagery; MR: Mental Rotation, MC: Mental Calculation (subtraction)
* Personal participation in the data acquisition
** Implementation of the study and data acquisition for this PhD thesis
*** Implementation and data acquisition before the beginning of this PhD thesis (2015-2016)

Figure 4.2: Summary table of datasets used in our different studies.

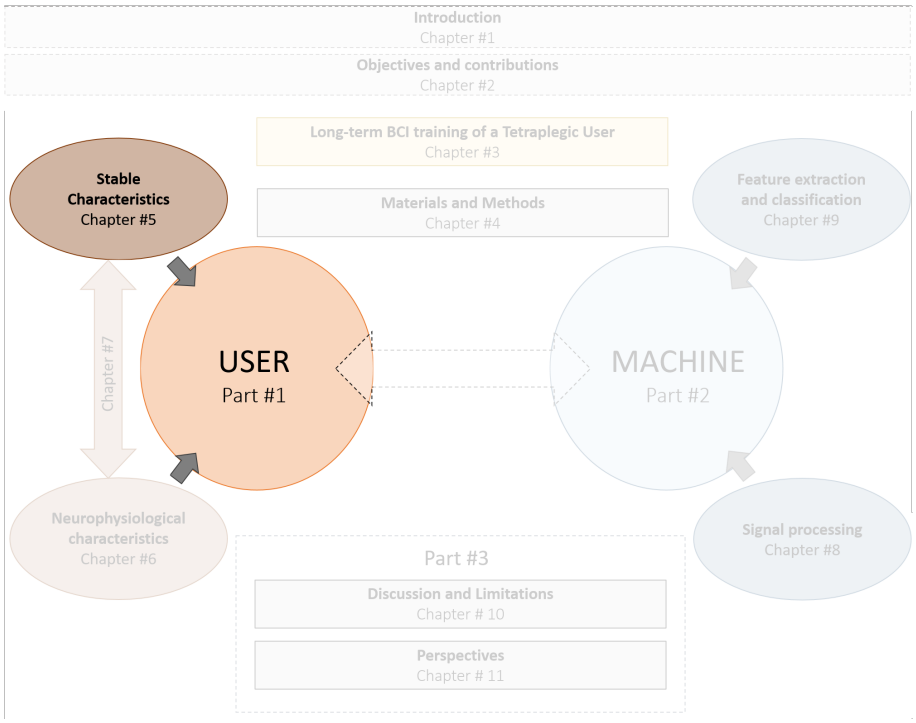
Part I

User-related Characteristics

5

Stable Characteristics

ROADMAP -



QUICK SUMMARY -

In the following chapter, we aim to understand the mechanisms underlying MT-BCI user-training, notably by identifying stable user-related factors (i.e traits) influencing it. Ideally, such factors of interest should generalize across experiments and datasets. So far, however, most factors that have been identified as predictors of BCI performance were studied in single experiments and/or data sets. Therefore, in this chapter, we study whether statistical models can predict and explain MT-BCI user performance across experiments, based either on:

- user traits (e.g., personality or gender) for between-subject predictions.
- user traits and information regarding their previous BCI performances for within-subject predictions.

Our models rely on 43 subjects, for a total of 180 MT-BCI sessions collected from three different studies based on the same MT-BCI paradigm. We used machine learning regressions with a leave-one-subject-out cross validation to build both between-subject and within-subject predictive models of performance. Our between-subject models suggested that whereas users' traits could predict between-experiment performances for some data sets, they may not be informative enough to reliably predict MT-BCI performances across experiments with few sessions and high between- and within-subject variability. However, we uncovered a within-subject model of MT-BCI performance that could predict the performances obtained at session N+1 based on both the MT-BCI performances of session N and anxiety-related traits.

Contents

RELATED PAPER/CONFERENCE PAPER/CONFERENCE ABSTRACTS -

BENAROCH C. ,JEUNET C., and LOTTE F., Can a computational model predict Mental-Task BCI performance across experiments based on users' characteristics", *Submitted in PLOS ONE*.

BENAROCH C. ,JEUNET C., and LOTTE F., (2019). Are users' traits informative enough to predict/explain their mental-imagery based BCI performances?. *In 8th Graz BCI Conference 2019*.

BENAROCH C. ,JEUNET C., and LOTTE F., (2019). Computational modelling to predict/explain MI-BCI users' performances and their progression. *In Journées Cortico-COllectif pour la Recherche Transdisciplinaire sur les Interfaces Cerveau-Ordinateur*.

5.1 Review of the literature

Important between- and within-subject performance variability has been observed in most of MT-BCI papers. This has led the community to look for predictors of MT-BCI performance and more particularly individual stable characteristics that could explain these variabilities. Indeed, identifying such predictors, prior to any BCI training, would allow BCI designers to implement the most suitable BCI protocol for a given user. Alternatively (or additionally) those predictors could improve our understanding about why some users fail to control a BCI and propose alternative training approaches for those users. This introduction offers a review of the latest developments in our understanding of the demographic, psychological and cognitive factors reported to influence MT-BCI performance (see Table 5.1).

Concerning socio-demographic characteristics, in [73] a positive interaction has been revealed between the subjects' age, amount of daily hand-and-arm movements corresponding to the practice of video games, musical instruments or sports, and their mu-power at rest, which itself has been shown to correlate with MT-BCI performances [14]. Moreover, Randolph et al., [67] have suggested that playing at least one instrument, not being on effective drugs, being a woman, and being over the age of 25 increased the likeliness of obtaining high MT-BCI performances.

Beyond socio-demographic variables, psychological traits like self-reliance and apprehension have been shown to linearly correlate with MT-BCI performances [52], just as mental rotation scores do, suggesting that spatial abilities influence MT-BCI performances. This last correlation has been replicated in two further studies [8, 65]. However, recently, Leeuwis et al. [142, 143] investigated the impact of spatial ability and visuospatial memory on MI-BCI performances. In their experiment, they did not confirm the impact of spatial ability on MI-BCI performance but reported a significantly higher visuospatial memory in BCI users with high performances. Finally, a significant positive correlation between BCI performances and visuo-motor coordination abilities has been revealed in [61] and replicated in [144], strengthening the fact that spatial abilities might strongly influence MT-BCI users' performances. Once the factors influencing MT-BCI performance identified, this influence can be quantified using modeling. For instance, Jeunet et al [52] experimentally revealed a model including four factors (mental rotation scores, self-reliance, apprehension and the visual/verbal sub-scale of the Learning Style), using a step-wise linear regression. The average prediction error of this model was lower than 3%. Using the same methodology (i.e., step-wise linear regression), Leeuwis et al. [143] linear models were able to predict the BCI performances (assessed in term of error rate) based on factors such as gender, vividness of visual imagery, personality traits and emotional stability. In their study, the best results were achieved by a model based on gender, emotional stability, orderliness and vividness of visual imagery. None of the models exceeded 40% explained variance and they were not used to predict performances of unseen users. Finally, Hammer et al [61] designed a model including the visuo-motor coordination factor and tested it across different studies [63]. The average prediction error of this model was below 10% for more than 50% of the subjects.

To be useful, models of prediction of MT-BCI performance should be stable and accurate, i.e., be able to actually predict the BCI performances of new users. These models should also consider multiple variables and should generalize across experiments and datasets. Unfortunately, it is not the case of most of the predictors found in those studies (i.e., see Table 5.1). Explanations could come from the fact that, too often, predictors of MT-BCI performances are found on experiments using either a low number of subjects or only one MT-BCI session. It may also be due to a large pool of potential predictors increasing the chances of detecting one. Indeed, when using a low number of subjects or sessions, within-subject variability could play a large role in the obtained performances as subjects performances vary over the course of individual session and their skill in operating a BCI does not remain constant over the course of a session [78]. In addition, in many studies [52, 61, 63, 67] a minority of predictors are found among the proposed ones.

Table 5.1: Studies concerning predictors from stable characteristics of MI-BCI performance.

Study	Number of subjects	Number of BCI sessions	Tasks	Tests	Predictors
Burde and Blankertz, 2006	17	1	Left Hand Right Hand Right Foot	• Locus of control of reinforcement (LOC)	✓ Locus of control for dealing with technology (+)
Randolph et. al, 2010	55	1	Left Hand Right Hand	• Age • Time spent typing per day • Time spent on activities requiring hand-and-arm movement per day • Time spent on activities requiring most of the body per day	✓ Interaction of age and daily amount of hand-and-arm movement (+)
Randolph 2012	80	1	Hand Foot	• Age (under/over 25) • Gender • Plays instrument (yes/no) • Plays sports (number of sports) • Video game experience (high/low) • Affective drugs (yes/no)	✓ Video game experience (+) ✓ Playing at least one instrument, not being on effective drugs, being a woman, and being over the age of 25 (+)
Hammer et. al, 2012 Hammer et. al, 2014	80 32	1 1	Left Hand Right Hand Right Foot	• Two-hand coordination (2HAND) • Cognitrone (COG) • Raven's standard progressive matrices (SPM) • Verbal learning test (VLT) • Non-verbal learning test (NVLT) • Big five plus one personality-inventory (B5PO) • Fragebogen zu Kontrollüberzeugungen (IPC-Scales) • Achievement motivation-test • Questionnaire for current motivation • Attitudes towards work (AHA) • Barrat Impulsiveness Scale (BIS-15) – *only in Hammer 2014* • Allgemeine Depressionsskala – ADS-L • Symptom Checklist-90-Revised – SCL90-R • Current mood	✓ Visuomotor coordination ability – 2HAND (+) ✓ Ability to concentrate on the task - AHA (+) → X Not replicated in Hammer 2014 ✓ Attentional impulsivity (-) - **BIS-15**
Cassady et. al, 2014	36	3	Left Hand Right Hand Both Hands Rest	• Mind-body awareness training (MBAT) – at least one year of experience, 2 times/week during at least one hour vs control group	✓ MBAT training (+)
Jeunet et. al, 2015	18	6	Left Hand Mental rotation Mental subtraction	• 6 subscales of the Wechsler Adult Intelligence Scale • Corsi Block task • Revised Visual retention test • Learning Style Inventory (LSI) • 16 Personality Factors—5 (16pf5) • Internal, Powerful others and Chance scale (IPC) • State Trait Anxiety Inventory (STAI) • Bruininks-Oseretsky Test of Motor Proficiency (BOT-2) • Mental Rotation test (MR) • Arithmetic test (WAIS-IV dimension)	✓ Apprehension (+) ✓ Self-reliance (+) ✓ Visual/verbal – LSI ((+) ✓ Spatial abilities (MR) (+) } 16pf5 → All combined in a linear model
Jeunet et. al, 2016	20	1	Left Hand Right Hand	• Mental Rotation test (MR)	✓ Spatial abilities (MR) (+)
Pacheco et. al, 2017	7	1	Flexion or extension Imagination (arm)	• Perceptual Reasoning (WAIS-IV dimension) • Block Design • Matrix Reasoning (WAIS-IV dimension) • Visual Puzzles (WAIS-IV dimension) • Mental Rotation test (MR)	✓ Block design test (+) ✓ Matrix reasoning test (+) ✓ Visual Puzzles (+) ✓ Mental Rotation test (+)
Zapala et. al, 2020	40	1	Left Hand Right Hand	• Edinburg Handedness Inventory	✓ Handedness → $Perf_{Left\text{-}handers} < Perf_{Right\text{-}handers}$
Leeuwis et. al, 2020 Leeuwis et. al, 2021	55	1	Left Hand Right Hand	• The Demographic questionnaire • The Affinity for Technology Interaction Scale (ATI) • The Vividness of Visual Imagery Questionnaire (VVIQ) • The Five Factor Personality Inventory (FFPI) • The Questionnaire for Current Motivation (QCM) • Mental Rotation test (MR) • Design organization test (DOT)	✓ Gender (Female>male) ✓ Emotional Stability (+) ✓ Orderliness (-) ✓ Vividness of Visual Imagery (+) → **All combined in a linear model** ✓ **Visuospatial ability (DOT) (+)** X **Spatial abilities (MR)**

(+) Positive relationship
(-) Negative relationship

5.2 Objectives

The review of literature introduced in Section 5.1 shows that several studies reveal relationships between BCI performances and stable characteristics of the user. While potentially insightful, these results present some limitations: they have been revealed in single experiments, each including a small number of subjects and sessions; or they correspond to univariate models (i.e., including a single factor). Therefore, in this chapter by combining data from three different experiments based on the same BCI paradigm, we studied the feasibility of determining stable, accurate and generalizable multivariate models that would explain/predict MT-BCI performance variability. Our question was: can we build a computational model that explain/predict MT-BCI performance across experiments based on users' characteristics? Indeed, through the factors included in the model, we would be able to better understand the between- and within-subjects variability and a computational model would allow us to predict performances of unseen participants. Therefore, combining both, an appropriate training could be propose to new users'. The subjects of the included datasets took part in three (for two of the dataset) or six (for one dataset) MT-BCI sessions, each session being divided into five runs. We gathered data from 43 subjects, for 180 BCI sessions in total. In these three experiments, the subjects had to complete psychometric tests and were asked to learn to perform three MT tasks, namely, left-hand movement imagination, mental rotation and mental subtraction. We created six groups from the three datasets in order to pair the subjects of the different experiments according to their specific experimental paradigms. We used a LASSO (Least Absolute Shrinkage and Selection Operator) [125] regression to determine explanatory and predictive models of within and between-subject MT-BCI performance variations, for each group. Within-subject MT-BCI performance variations stand for the performance variations of one subject across different sessions and between-subject MT-BCI performance variations for the performance variations inside one group of subjects. We used features describing their gender, age, personality and cognitive profile to build those models. We chose the LASSO as it promotes sparse solutions and thus allows us to have interpretable models.

5.3 Materials and methods

5.3.1 Datasets

To maximize the number of subjects, we used data from three different experiments [52, 64, 132, 133]: Datasets 3A, 3B and 3C. They were all based on the same BCI paradigm, i.e., MT-BCI with three mental tasks: left-hand motor imagery, mental rotation and mental subtraction. The materials and methods used are presented in Figure 5.1 and Figure 5.2 and additional information is available in Chapter4.

As the three studies had different numbers of sessions, we made 7 groups out of the 3 datasets. The last group i.e, *Group 7* is only used for the within-subjects analyses. We chose those groups in a way that each group had comparable sessions. The first three ones correspond to the BCI sessions of the three datasets taken separately (i.e., 3A, 3B and 3C). The fourth group corresponds to sessions 1, 5 and 6 of both 3A and 3B, since sessions 2, 3 and 4 of 3B were not BCI sessions but SA or VC (Verbal Comprehension) sessions, see Figure 5.2. The fifth group gathers sessions 1, 2 and 3 of 3A and the three sessions of 3C. *Group 6* corresponds to all the 43 subjects together and includes the first session of 3A, 3B and 3C as it is the only session where all subjects were trained to BCI control at the same time in the protocol and following the exact same conditions (see Figure 5.3). For the within-subject analyses, *Group 7* included the first two BCI sessions of each dataset, i.e., sessions 1 and 2 of 3A and 3C and sessions 1 and 5 for 3B.

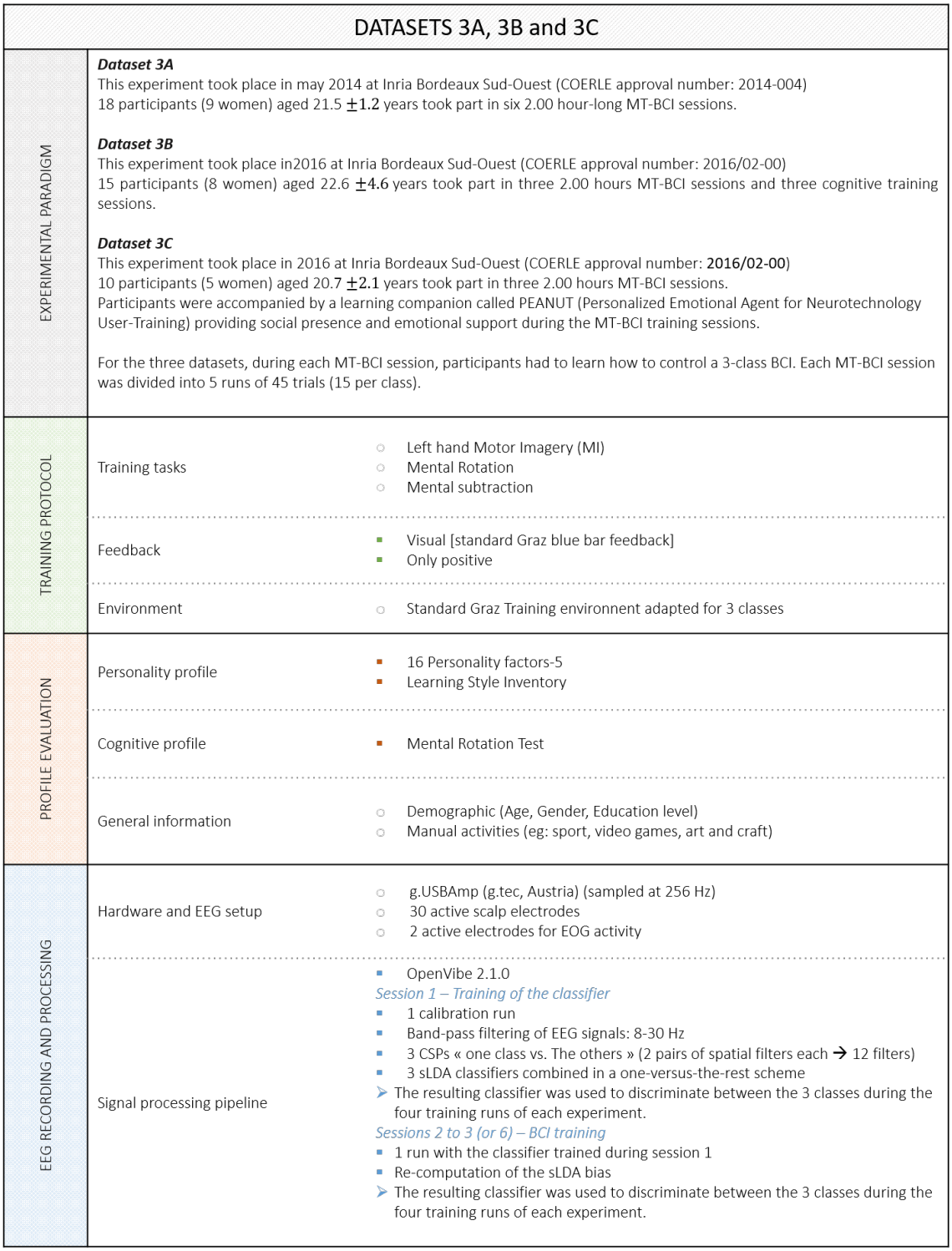


Figure 5.1: Summary table for Datasets 3A, 3B and 3C.

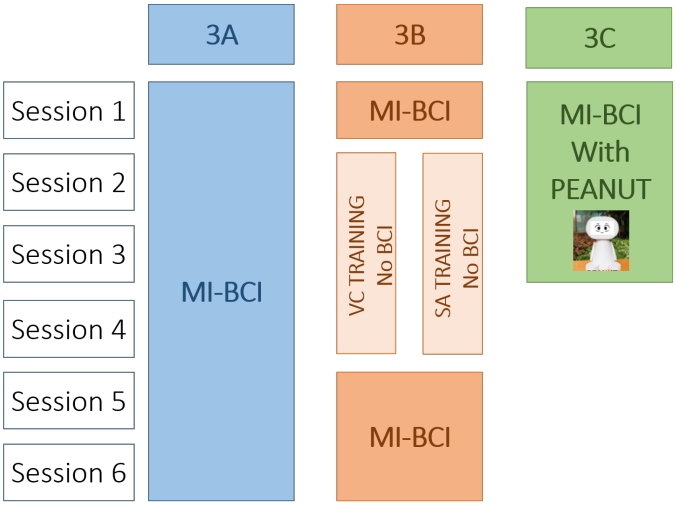


Figure 5.2: Details of the three different studies: 3A, 3B, 3C. In 3C, PEANUT (Personalized Emotional Agent for Neurotechnology User-Training) is a learning companion.

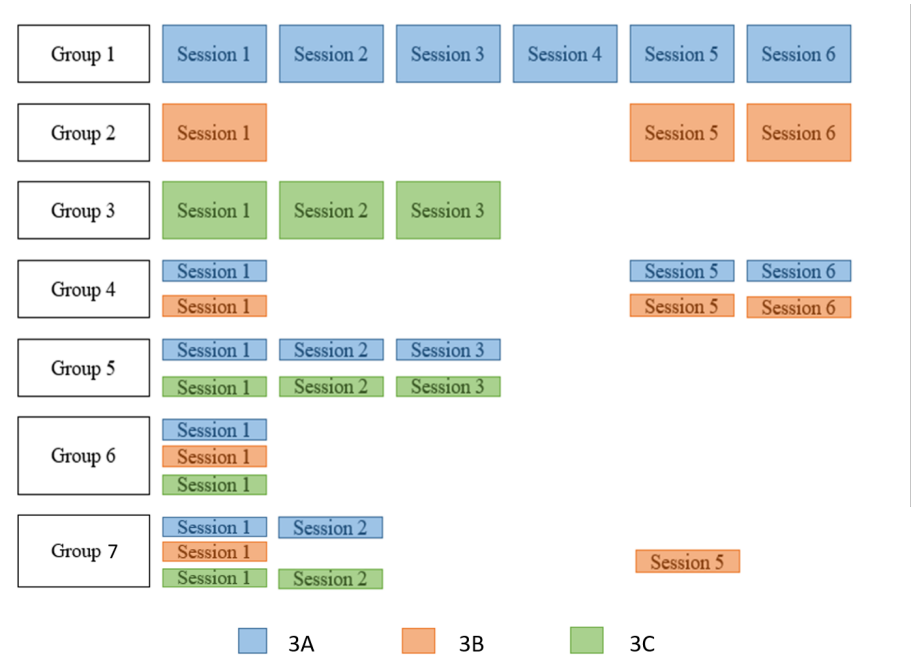


Figure 5.3: Details of BCI sessions for each group. In blue sessions concerning 3A, in orange 3B and in green 3C

5.3.2 Variable to predict: BCI Performances

MT-BCI performance was assessed in terms of mean classification accuracy $Perf$, i.e., the mean accuracy measured over all the windows of the feedback periods of the runs 2 to 5, i.e., all 1s long sliding windows -separated by 0.0625s (with overlap)- between t=4.250s and t=8.250s of each trial. We used the mean classification accuracy $Perf_{real}$ over the different sessions as the target variable to be explained/predicted by our models. For each experiment, all subjects managed to control the BCI interface better than chance, on average across all sessions. Indeed, their average classification accuracies (across all sessions) were higher than chance level, which was estimated at 37.7% of correct classification regarding the following characteristics: three classes, more than 160 trials per class and $\alpha = 0.05$ [122]. However, when considering only one session (i.e., 60 trials per class), four subjects were under the chance level, which was then estimated at 40.1% of correct classification for 60 trials per class and $\alpha = 0.05$. The predictive models provided us with a predicted performance $Perf_{pred}$. Our objective consisted in minimizing the mean absolute error ($|Perf_{real} - Perf_{pred}|$).

5.3.3 Explanatory features

5.3.3.1 Between-subject prediction:

In the 3 studies, subjects were asked to complete psychometric and personality questionnaires, which aimed to assess different aspects of their personality and cognitive profile. A series of questionnaires had been completed by the participants in each experiment. For this study, we considered only the questionnaires that had been used in all the three experiments. The 16 Personality Factors 5 (16PF5) provided a score for sixteen primary factors of personality and the Mental Rotation test was used to assess the subjects' spatial abilities i.e., their ability to produce, manipulate and transform mental images (see Chapter 4, Section 4.2)

In total, we used 17 parameters to represent the personality and cognitive profiles of each participant. We also added information about the age and the gender (binary variable i.e., 1 for male and 2 for female) of the subjects. We thus had 19 features (i.e., traits) available to find a predictive/explanatory model for each experiment. When combining several studies together, we also added information about the ID of the experiment (using vectorization i.e., one feature for each study with a value of 1 when the subject took part in this specific study and 0 otherwise). Thus, 21 or 22 features were available to find a predictive/explanatory model when combining two or three studies, respectively.

5.3.3.2 Within-subject prediction:

We also aimed to predict the progression of BCI users in terms of classification accuracy (i.e., performance) evolution across sessions. More precisely, we attempted to predict the performances that participants would obtain at a session N+1 based on: the performances they obtained during session N, the time lapse between both sessions (i.e., the number of days between session N and session N+1 of BCI training), their traits (the same as for the Between-subject prediction) and, when relevant, the ID of the study. In this chapter, we will study the evolution between session 1 and 2 of BCI training (i.e., N=1).

5.3.4 Analyses

In order to build a Statistical model to predict/explain MI-BCI performances we used a LASSO regression to obtain models that could predict the performances of MT-BCI users from the stable characteristics of the user (i.e., traits). See Table 5.2. (see Chapter 4, Section 4.6 for more details about LASSO regression and methodology).

Table 5.2: Summary table explaining the statistical model used to predict/explain MT-BCI performances.

STATISTICAL MODEL OF PREDICTION		
METHOD	LASSO	
EXPLANATORY FEATURES	Between-subjects prediction	Within-subjects prediction
	▪ 16 personality factors (16pf5) ▪ Mental Rotation score ▪ Age ▪ Gender	▪ 16 personality factors (16pf5) ▪ Mental Rotation score ▪ Age ▪ Gender ▪ Number of days between the first two MT-BCI sessions ▪ MT-BCI classification accuracy at session 1
VRIABLE TO PREDICT	MT-BCI mean classification accuracy	
EVALUATION OF STABILITY AND RELIABILITY	▪ Leave-one-out cross validations ▪ Random model with 1000 permutations	

5.4 Results

5.4.1 Results of between-subject predictions

5.4.1.1 Outliers detection

We excluded from the analyses the subjects whose mean classification accuracy was above or below two standard deviations (SD) from the group's mean performance (see summary in Table 5.3).

Table 5.3: Group details

Group	$Perf_{Group}$	SD	Outliers (Mean Performance)	Final number of subjects (Between-subject analyses)	Final number of subjects (Within-subject analyses)
1	52.50%	5.62	Subject 1 (67.21%)	17/18	18/18
2	50.64%	9.47	-	15/15	15/15
3	50.74%	7.77	-	10/10	10/10
4	51.48%	7.87	-	33/33	33/33
5	52.04%	6.40	Subject 1 (66.00%) Subject 37 (38.97%) Subject 38 (38.49%)	25/28	28/28
6	53.27%	9.54	Subject 7 (32.80%)	42/43	43/43

5.4.1.2 Models of prediction

The results indicated that before adjusting the p-values for multiple comparisons, only the predictions made for *Group 1* were better than chance ($p = 0.02$), with a mean absolute error of 2.98% (see Table 5.4). For this group, 80% of the generated models included the same three factors: the Mental Rotation score, the Self-reliance score and the Tension score with a negative weight for the Tension score (see Figure 5.4, *Group 1*). For more details on these factors see Appendix A. After adjusting the p-values for multiple comparisons using Benjamini-Hochberg (BH) procedure [145], the prediction performance was better than chance for none of the groups. For all groups, the selected factors were stable across the cross-validations. However, when the between-subjects performances showed a high variability, our models were not able to predict them (see Figure 5.4, *Group 3, 4, 6*).

The chance levels for each group are displayed in Table 5.4. We also computed the correlation between the real and predicted MT-BCI performances for each subject. We only obtained a significant correlation for *Group 1* [$r = 0.62$, $p = 0.06$, $p_{adj} = 0.03$].

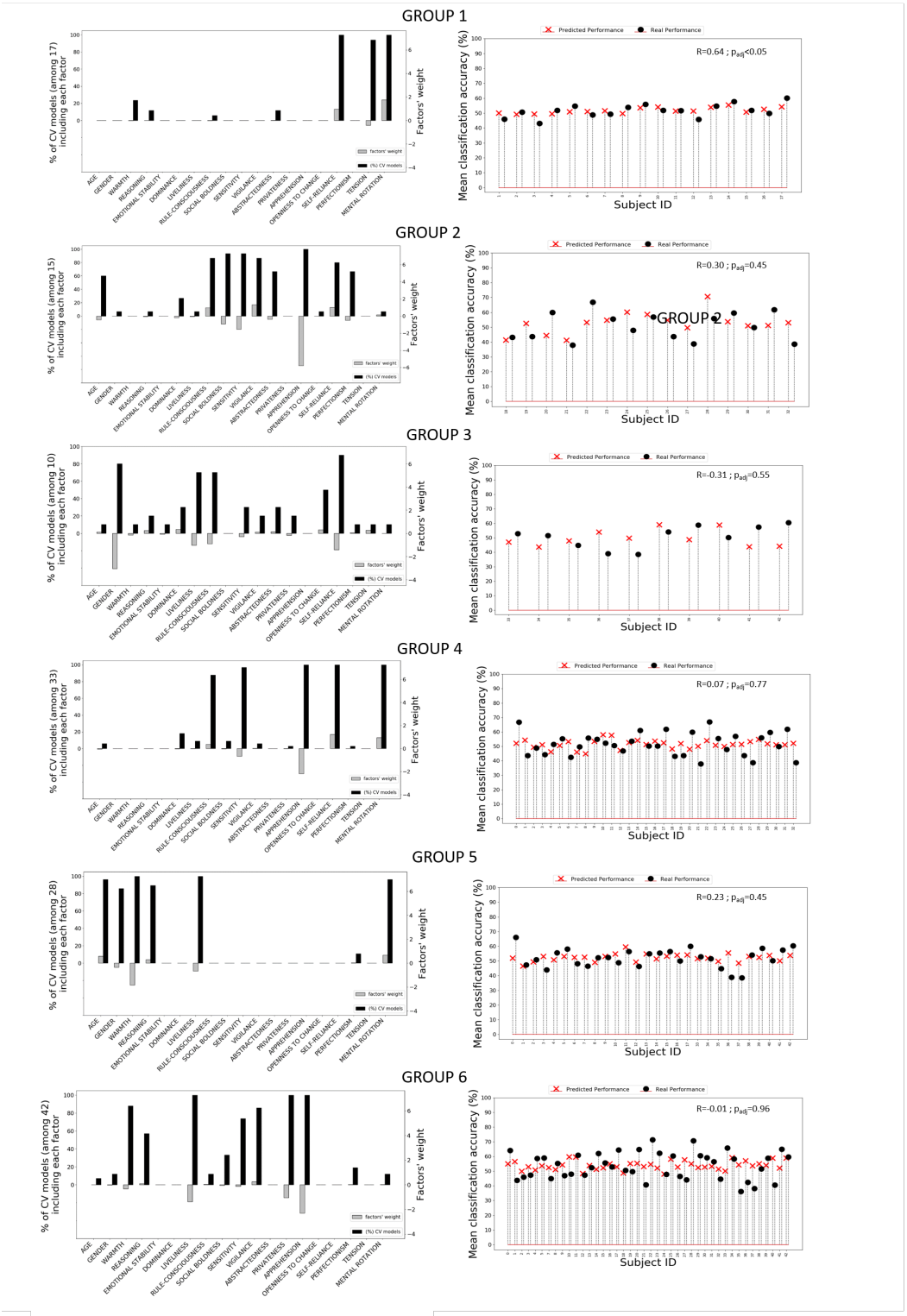


Figure 5.4: Results of the different between-subjects models generated for each Groups. On the left, the percentage of Cross-Validation models including each factor (black) and their weight (grey). On the right, in black (circle), the real performance of each subject and in red (cross), the predicted performance of each subject generated using the model generated from the training dataset (All subjects except the target one). Finally, in the right plots, the correlation between the real and predicted performances.

Table 5.4: Comparison of the mean absolute error with the mean absolute error of the random model (after 1000 permutations) for the between-subject prediction

Group	Mean absolute error (%)	pValue	adj pValue	Mean absolute error (%) of random model ($p < .01$)	Mean absolute error (%) of random model ($p < .05$)
1	2.92	0.02	0.09	2.67	3.31
2	8.46	0.15	0.25	5.37	6.69
3	9.66	0.87	0.87	5.14	5.64
4	6.80	0.78	0.87	6.19	6.34
5	4.87	0.10	0.24	4.30	4.70
6	7.88	0.23	0.33	6.90	7.38

5.4.2 Intermediate discussion

Between-subject analyses suggested that it was not possible to predict the mean MT-performance of BCI users although, for *Group 1*, our model was better than chance before adjusting p-values ($p = 0.02$ $p_{adj} = 0.09$). The prediction models of *Group 1* included three main factors: The self-reliance, the tension and the mental rotation scores, i.e., similar factors to the ones found in Jeunet et al [52] using a step-wise-linear-regression. Indeed, the self-reliance, apprehension, visual/verbal subscale of the learning style inventory [146] and the mental rotation scores were included in their model. It should be noted that both the apprehension and tension factors, included in our models with a negative weight, are related to the same global factor, anxiety. The anxiety factor represents the extent to which people can control their emotions and reactions. High-scorers may find it difficult to control their emotions or reactions and therefore act in counterproductive ways whereas low-scorers tend to be unperturbed. Moreover, the self-reliance factor was also selected in 100% of the CV models of *Groups 2, 4 and 6* with a large weight, even though the prediction reliability for these models was not better than chance, This highlights its apparent strong influence.

Furthermore we can observe that the mean absolute error of *Group 3* (see Table 5.4) was far larger than the one obtained with the random model. As mentioned before, this group comprised only 10 subjects who performed 3 BCI sessions. Consequently, the between-subject variability may have been too high to find stable models of prediction. Indeed, 17 features were included in the cross validated models in total and only two (i.e. the gender and self-reliance) were in 80% of them (see Figure 5.4) with a negative weight of -3.04±1.76 and -1.42±0.9 respectively.

Interestingly enough, the experiment during which datasets 3B and 3C were collected had been designed to influence the factors that had been identified through 3A (used in Jeunet et al [52]) as potentially influencing MT-BCI performance: with a spatial ability training in 3B dedicated to the improvement of the participants' mental rotation abilities, and a learning companion in 3C aiming to help the anxious and non self-reliant subjects. Therefore, it makes sense to observe a reduced influence of the mental rotation scores and apprehension/self-reliance factors in the groups including the data of 3B and 3C, respectively. Indeed, the self-reliance factors assesses the tendency to seek companionship and enjoy belonging to and functioning in a group (inclusive, cooperative, good follower, willing to compromise). High-scorers tend to be more individualistic and self-reliant and to value their autonomy.Therefore,one could hypothesize that using a learning-companion could negatively impact performances of self-reliance high-scorers.

5.4.3 Results of within-subject predictions

5.4.3.1 Outliers detection

No outliers were detected for the within-subject prediction.

5.4.3.2 Models of prediction

Our results for the within-subjects prediction, i.e., the prediction of performances obtained at session N+1 according to those of session N, revealed a model of prediction better than chance ($p = 0.001$) for *Group 2* (Dataset 2B) and *Group 7* (Dataset 2A, 2B and 2C)((see Table 5.5). For those two groups, the performance obtained by the subject at session N was selected in 100% of the generated models with a very strong weight in comparison to the other factors' weights (see Figure 5.5). The vigilance score (measuring the tendency to trust others' motives and intentions) was also selected in 80% and 100% of the models for *Group 2* and *Group 7* respectively (see Figure 5.5) with a positive weight of 2.3±0.5 and 2.3±0.2 respectively. We also computed the correlation between the real and predicted MT-BCI performances for each subject. We obtained a significant correlation for *Group 2* $r = 0.79$ ($p \leq 0.001$, $p_{adj} < 10^3$) and *Group 7* $r = 0.59$ ($p \leq 0.001$, $p_{adj} \leq 0.01$) (see Figure 5.5).

In addition, we computed the correlation between online performances of session N and online performances of session N+1 (see Figure 5.6). Results show high correlation $r = 0.87$ ($p_{adj} = 3.10^{-4}$) and $r = 0.63$ ($p_{adj} = 2.10^{-5}$) for *Group 2* and *Group 7* respectively.

Table 5.5: Comparison of the Mean absolute error with the mean absolute error of the random model (after 1000 permutations) for the Within-subject prediction

Group	Mean absolute error (%)	pValue	adj pValue	Mean absolute error (%) of random model ($p < .01$)	Mean absolute error (%) of random model ($p < .05$)
1	8.04	0.52	0.65	5.70	6.140
2	5.22	0.001	0.005	6.25	7.48
3	6.53	0.12	0.24	4.29	5.35
7	6.12	0.001	0.005	6.77	7.47

5.4.4 Intermediate discussion

The results for the within-subject models (see Figure 5.6) suggested that the mean classification accuracy of session N (here N=1) could allow us to predict the one of session N+1 for two group (i.e., *Group 2 and 7*). However, for those two groups, the performance of session N was included in the models with a far higher weight than other included factors (-7.44±0.8 and 4.11±0.2 respectively). This result could be attributed to the strong correlation between online performances of sessions N and online performances of session N+1, suggesting a small increase in performances between those two sessions.

In addition, in models generated using *Group 2 and 7*, three factors included in the global factor anxiety were selected in at least 80% of the cross-validated models: the emotional-stability factor (with a negative weight of -0.63±0.42 and -0.53±0.20 for *Group 2 and 7* respectively), the vigilance factor with a positive weight of 2.32±0.47 and 1.53±0.19 for *Group 2 and 7* respectively and the apprehension factor with a negative weight of -1.31±0.60 and -0.93±0.24 for *Group 2 and 7* respectively. Anxiety includes tendencies to be reactive (emotional-stability -) rather than able to adapt, vigilant and suspicious (vigilance +), worried and self-doubting (apprehension +) and tense (tension +). People with little anxiety tend to be more unperturbed. In the context of BCIs and within-subject variability, our models would suggest that users' who can adapt, anticipate and who are vigilant and confident have higher future performances.

In addition, two other factors were selected in at least 80% of the cross-validated models: the dominance factor (with a negative weight of -0.01±0.04 and -1±0.18 for *Group 2 and 7* respectively) and the sensitivity factor (with a negative weight of -0.007±0.03 and -0.47±0.19 for *Group 2 and 7*

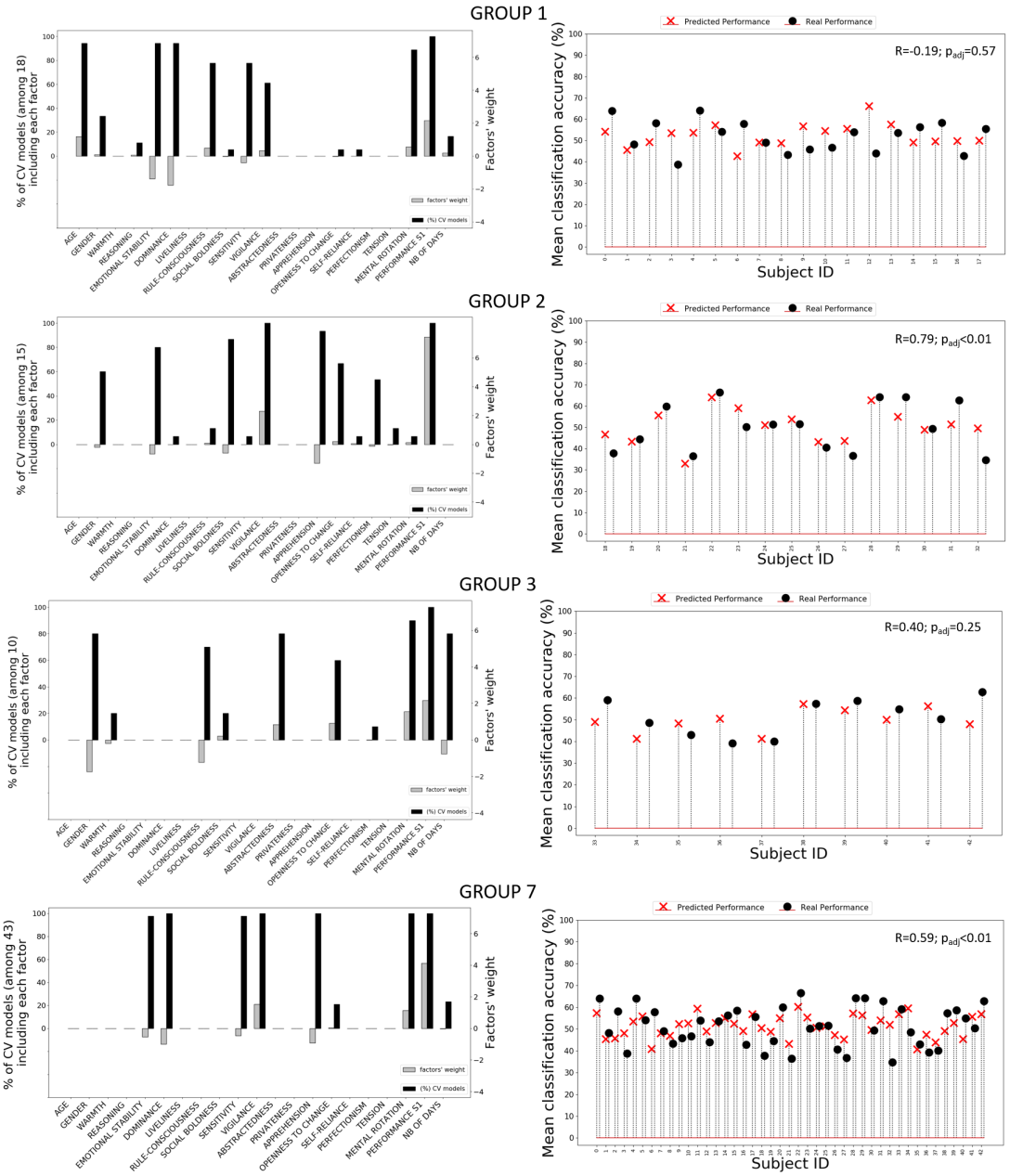


Figure 5.5: Results of the different within-subjects models generated for each Groups. On the left, the percentage of leave-one-out cross-validation models including each factor (black) and their weight (grey). On the right, in black (circle), the real performance of each subject and in red (cross), the predicted performance of each subject generated using the model defined from the training dataset (All subjects except the target one). Finally, in the right plots, the correlation between the real and predicted performances.

respectively). A high score for the dominance factor might be a sign that a person is being determined and combative when they want something. Sensitive subjects (i.e., high-scorers for sensitivity factor) rely on empathy and affect and are less oriented towards the utilitarian aspect of things and how they can work. Using those definition, our models would suggest that determined utilitarian and objective users have higher future performances.

Finally, the reasons why only *Groups 2 and 7* yielded significant results remain unclear although the high correlation between online performances of session N and N+1 might be one explanation. Another possibility could be that the between-subject variability was higher in those groups, which can make the prediction easier.

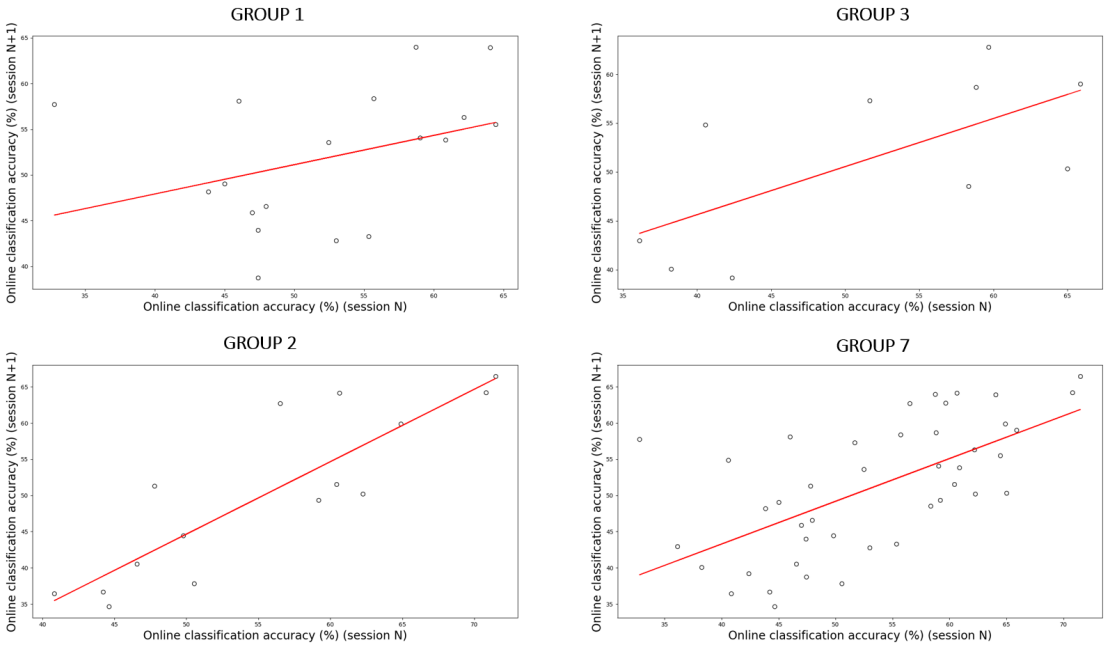


Figure 5.6: Regression analysis between online performances of session N and online performances of session N+1.

5.5 Discussion

In this study, we gathered the data of 3 experiments to study the feasibility of predicting/explaining MT-BCI performances across experiments, using a statistical model based on the subjects' traits only. The objective was, through those models of prediction, to identify stable factors across experiments, that are associated to subject performances. On the longer-term, we aim to adapt the training of protocols based on those factors in order to optimize the learning. In order to have interpretable models we used the LASSO, as this method promotes sparse solutions. Our analyses suggest that the users' traits alone may not be sufficient to predict MT-BCI performances across experiments, at least with data collected from a small number (here 3 or less) of sessions. As any result, this negative result is nonetheless useful and necessary to better understand and model BCI user training, and notably to identify the factor that have an influence and those that do not [147].

One observation we can make concerns the many differences existing between the three studies. Indeed, the total duration of each study was different, they did not have the same number of participants, nor the same number of MT-BCI sessions. First, concerning the duration, we observed that for 3B, the number of days between two consecutive sessions was higher than the one observed for 3A and 3C. This is due to the protocol of 3B in which 3 verbal comprehension or spatial ability training sessions were included between the first and the second MT-BCI session. No significant differences were observed between the mean BCI performances of 3B and the ones of 3A and 3C. However, it should be noted that 3B participants' performance seem to decrease along the sessions. The reason of this decrease remains unclear. This phenomenon could be due to the fact that there is an average of 33 days between them or because the same classifier was used for each session (classifier biases were re-calculated after the first run of all the sessions except the first one). Moreover, 18 subjects were included in 3A for 6 BCI sessions whereas in 3B and 3C there were fewer participants, who took part in only 3 BCI sessions. The number of MT-BCI sessions seems to have an impact on mean BCI performance over all sessions because of the between-session variability (due, e.g., to fatigue or motivation fluctuations). Indeed those variations can be very important and can become an issue when computing the mean performance over all the MT-BCI sessions. When considering the traits of the users, averaging performances over 6 sessions enabled us to smoothen the performance variability, which made the computation of the mean MT-BCI performance more accurate and thus

allowed us to estimate more precisely the participants' actual ability to control an MT-BCI.

As traits are supposed to remain stable in time, having a more stable measure of performance (here averaged over more sessions) could help us find a more reliable model. In addition, for the three studies, the number of participants was quite low (not higher than 18) as well as the number of sessions (3B and 3C had only 3 sessions). When studying human learning, a larger number of subjects should be trained for a larger number of sessions. Indeed, the between- and within- subject variability make it hard to find stable models of prediction. Unfortunately, currently there is no large-scale BCI data bases including multiple subjects with multiple sessions per subject, with the same BCI system. Finally, stable characteristics were assessed using self-report personality test. Hence, results should be interpreted carefully as no interview was subsequently conducted to strengthen them. The fact that a large majority of the subjects were recruited among students from the science and technology campus of the local university and have a similar age could also induce a bias in the results. For instance, mental rotation scores can be influenced by the age or mathematical skills [128].

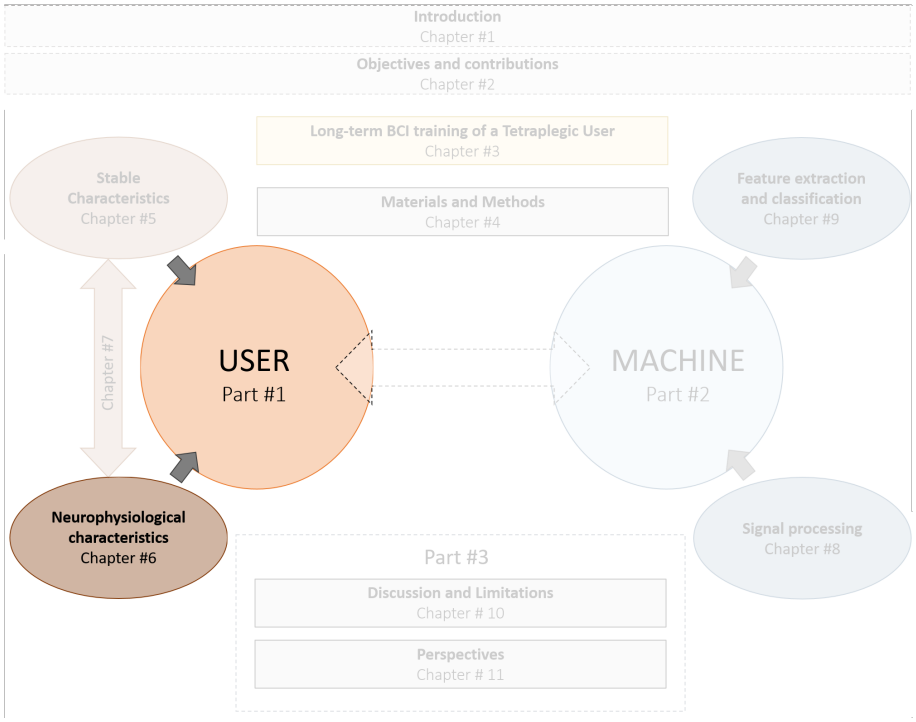
Regarding the statistical methods, we chose to use a LASSO regression because it appears to be more interpretable than a regular linear regression as it promotes sparse solutions. Moreover, the LASSO regression is usually more robust than a step-wise linear regression [148]. In a step-wise linear regression, choices are made locally at each step by fixing the weight, which is sub-optimal in general [149]. However, this method presents some limitations. The LASSO does not necessarily lead to a unique model. Indeed, if several explanatory variables are highly correlated, the LASSO may randomly chose one for its model, which might make the resulting model potentially unstable in this case [125]. An other method, i.e., the Elastic Net [141] regression can lead to a unique model. Nonetheless, it does not enable a variable selection as sparse and therefore as interpretable as with the LASSO. Therefore, we also explored this method, as a sanity check, in order to find predictive models of performances (see Appendix D). However, we obtained similar results as those obtained with the LASSO, and models were more complex, thus limiting their interpretability. This confirmed the relevance of the LASSO for this problem. Finally, by using the LASSO we hypothesized that there was a linear correlation between the included explanatory factors and MT-BCI performances. However, this correlation might also be non-linear, so it would be worth studying non-linear regression models in the future.

Finally, we could assume that to reliably predict MT-BCI performance, we would need larger samples and longer-term experiments. This would enable to maximally average-out inter-session variability and thereby, obtain a more stable mean performance across sessions that could be better related to users' traits. We hypothesise that by combining users' traits together with other factors such as neurophysiological predictors, feedback characteristics and the properties of the machine learning models learned from the data, we could be able to create a more robust and precise model of prediction for MT-BCI performances. Altogether, this section is a first step towards the computational modelling of MT-BCI performances.

6

Neurophysiological predictors

ROADMAP -



QUICK SUMMARY -

In the following chapter, we will try to understand the mechanisms underlying MT-BCI user-training, notably by identifying neurophysiological factors related/associated to it. Such factors of interest should generalize across datasets. However, so far, most neurophysiological factors that have been identified, individually, as factors related to BCI performance but not as predictors of performance of unseen users. Furthermore, they were studied in single experiments and/or data sets. Therefore, in this chapter:

- We compute and compare neurophysiological predictors of MI-BCI performance found in the literature using two of our Datasets i.e., 2A and 2B.
- We proposed new neurophysiological predictors of MI-BCI performance.
- We study whether statistical models can predict and explain MT-BCI user performance across experiments, based on users' neurophysiological characteristics.

We used data from 85 subjects, for a total of 340 online MI-BCI runs collected from two different studies based on the same MI-BCI paradigm. We used machine learning regressions with a leave-one-subject-out cross validation on one dataset (i.e. Dataset 2A) to build our predictive model of performance and, to study the reliability of the model, we used it to predict the performance of an other dataset (i.e. Dataset 2B). Our results suggested that it was possible to predict MI-BCI performances of unseen users using neurophysiological characteristics of a user with an error of 9.3% ($p = 0.005$). In addition, the proposed new predictors were found to be relevant predictors of MI-BCI performances in our analyses.

Contents

COLLABORATORS -

Eidan Tzdaka

RELATED CONFERENCE PAPER/CONFERENCE ABSTRACTS -

TZDAKA, E., BENAROCH C., JEUNET C., and LOTTE F., (2020). Assessing The Relevance Of Neurophysiological Patterns To Predict Motor Imagery-based BCI Users' Performance. *In 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 2490-2495). IEEE.*

TZDAKA, E., BENAROCH C., JEUNET C., and LOTTE F., (2020). Using neurophysiological predictors to predict MI-BCI users' performances. *In CORTICO COllectif pour la Recherche Transdisciplinaire sur les Interfaces Cerveau-Ordinateur.*

BENAROCH C., TZDAKA E., JEUNET C., and LOTTE F. Computational Modelling of MI-BCI performances based on users' neurophysiological characteristics. writing in progress.

6.1 Review of the literature

In the literature, different neurophysiological patterns were explored to identify predictors of MI-BCI performances. A part of them were based on different neurophysiological signals extracted using EEG. Therefore, in this section we will first present brain signals that can be measured through EEG and that can be used to drive a BCI. Then, we will propose a review of neurophysiological factors used to predict MI-BCI performances.

6.1.1 Brain signals and MI-BCI

EEG measures the electrical activity generated by the brain using electrodes placed on the scalp [150]. EEG measures aggregate electrical activity across millions of cells that have the same radial orientation with respect to the scalp. As EEG signals have a very weak amplitude, they are amplified before being digitized and processed. Electrodes are usually placed and named according to a standard model, namely the 10-20 international system [151]. This system has been initially designed for 19 electrodes, however, extended versions have been proposed to deal with a larger number of electrodes [152]. Those electrodes are placed in the scalp in order the measure electrical activity of specific brain regions (see Figure 6.1).

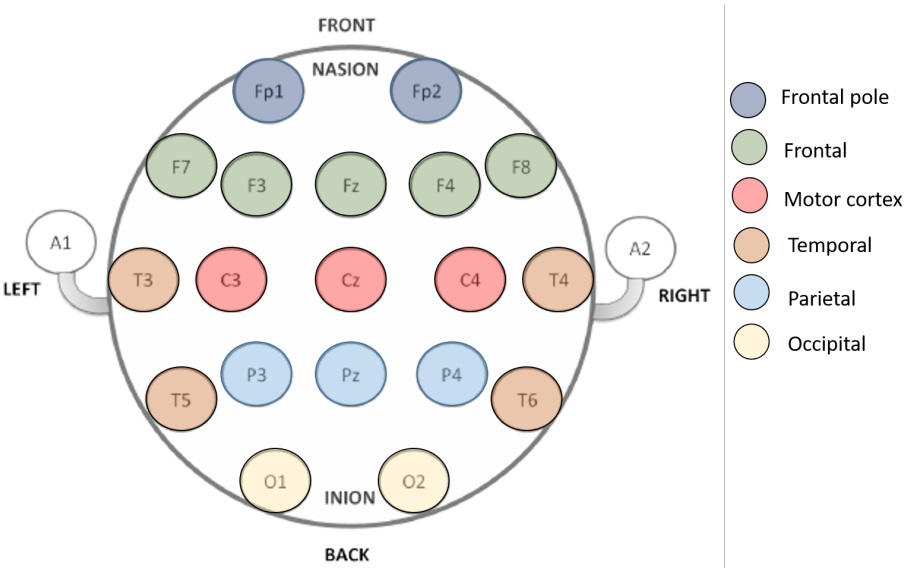


Figure 6.1: Positions, names and area of interest of the 10-20 international system electrodes.

EEG signals are composed of different oscillations named "rythms" [153]. Each rythms have distinct properties in terms of spatial and spectral localization and can be separated in 6 main classes (see Figure 6.2):

- **Delta (δ) rhythm** is a slow rhythm (1-4 Hz), with a relatively large amplitude. It is mainly found in adults during deep sleep [154].
- **Theta (θ) rhythm** is a slightly faster rhythm (4-8 Hz). It considered as important for different types of learning and memory [155, 156], as well as for synaptic plasticity [157, 158]. In addition, it promotes cognitive states such as anxiety [159]. Finally, it is well known to be present during states of deep relaxation such as stage 1 sleep, meditation and hypnosis [160] when the subject has closed-eyes. They are accessible across multiple brain areas.
- **Alpha (α) rhythm** are oscillations, located in the 8-12 Hz frequency band. They appear mainly in the posterior regions of the head (occipital lobe) when the subject has closed eyes or is in a relaxation state [161].

- **Mu (μ) rhythm** are oscillations in the 8-13 Hz frequency band, being located in the motor and sensorimotor cortex. The amplitude of this rhythm varies when the subject performs or imagine a movement [29].
- **Beta (β) rhythm** is a fast rhythm (12-30 Hz). It is observed in awaken and conscious persons [162] and is predominant in the parietal cortex [163]. In addition, it is affected, in the motor and sensorimotor cortex, by the performance of movements [29].
- **Gamma (γ) rhythm** is a rhythm which concerns mainly frequencies above 30 Hz. Different studies demonstrate that changes in amplitude or coherence of gamma-band oscillations relate to a broad range of processes, including feature integration [164] , stimulus selection [165], attention and memory formation [166], multisensory and sensorimotor integration, movement preparation [167–169], and even conscious awareness [164]. It is predominant in the frontal cortex [163].

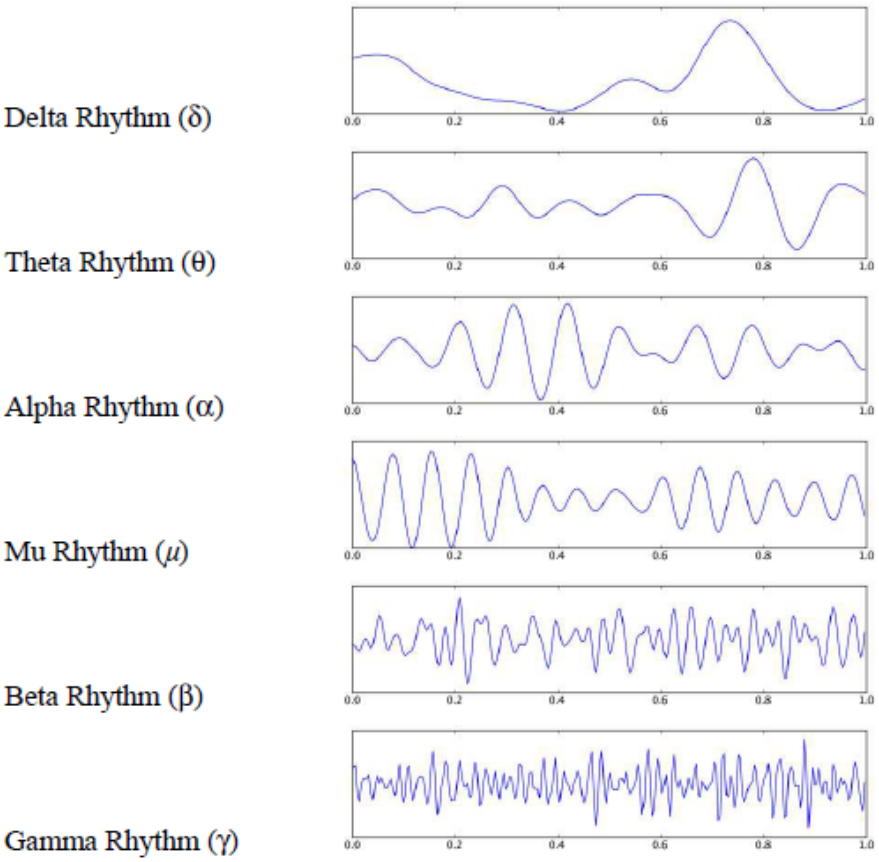


Figure 6.2: Different brain rhythms as measured by EEG. (pictures from wikipedia.org)

In MI-BCI, the aim is to identify, in the brain activity measurements of a given subject, one or several neurophysiological signals (i.e., brain activity patterns), in order to associate a command to each of these signals. More specifically, when performing motor imagery, a user has to imagine a movement of his own limbs (e.g., hands or feet). The signals resulting from such tasks are relatively easy to recognize as they have specific temporal, spectral and spatial features. Indeed sensorimotor rhythms (SMRs) [170], in the broad sense, refer to oscillations recorded over the sensorimotor cortex in the 8-30Hz frequency-band. An SMR desynchronisation, i.e., a decrease of signal power, is typically observed in the controlateral sensorimotor cortex during the execution, or imagination [171], of a hand movement [172]. On the same area, an increase of power is observed after completing a movement.

Those specific power decrease/increase that usually occur in the alpha or beta bands are referred to as event-related (de)synchronization (ERD/ERS) [173] (see Figure 6.3).

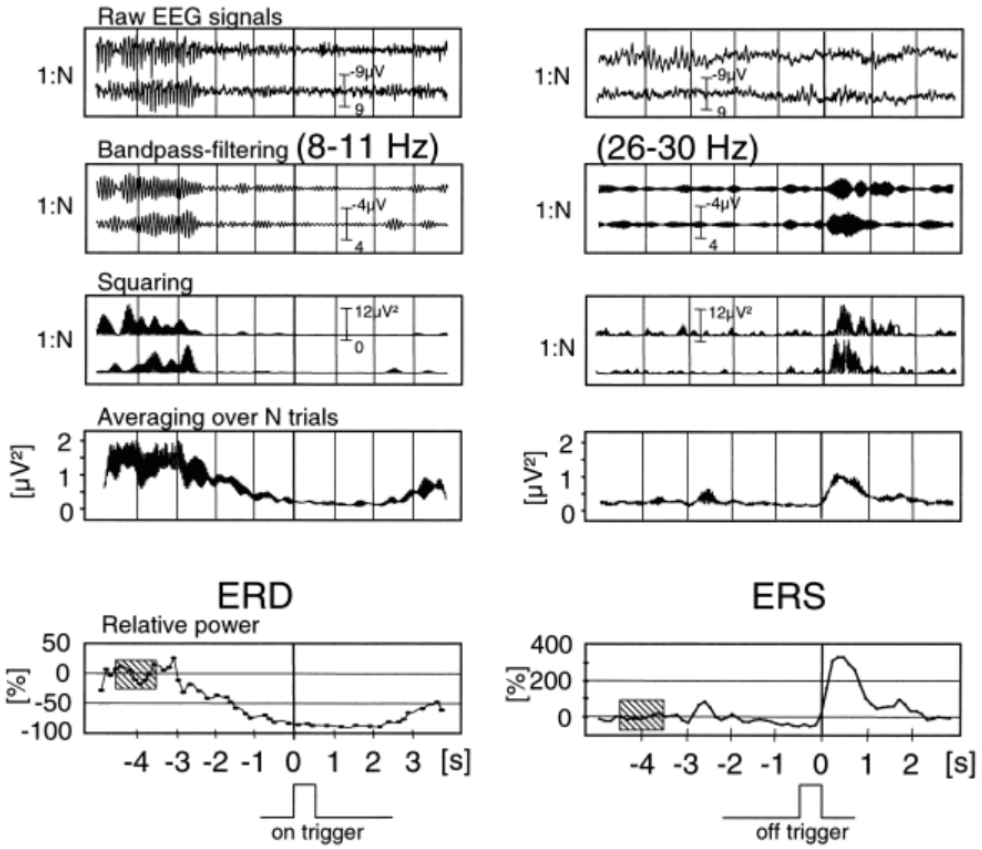


Figure 6.3: Principle of ERD (left panel) and ERS (right panel) processing. A decrease of band power indicates ERD and an increase of band power ERS. (Image from [173])

All those neurophysiological signals have been used in MI-BCI to find predictors of MI-BCI performances. The next section reviews the latest studies concerning neurophsyiological predictors of inter- and intra- subjects performance variabilities.

6.1.2 Neurophysiological predictors

Different neurophysiological predictors have been identify in the literature. Table 6.1 summarises those studies. A large majority of them are predictors of motor-imagery based BCIs (MI-BCIs). The main objective of those research is to evaluate a user's potential performance before conducting a time-consuming BCI experiment.

As motor learning processes and the human kinematic (i.e., mechanics of a living body) vary in each individual (i.e., speed, amplitude, trajectory among other factors are specific to each individual) focusing on EEG patterns associated with individual motor variability could help the BCI community to explain the variability in motor-imagery based BCIs. Indeed, when using an MI-BCI, users are asked to imagine a limb movements. Motor imagery generally leads to changes in brain rhythms originating from the sensory motor cortices, i.e., SMRs. Blankertz et., al [**?**] based their research on this phenomenon and used the users' ability to generate high power signals in the alpha band without performing any task. They proposed one of the most famous and reliable predictor of BCI performance (correlation of r= 0.53 $p < 0.05$ with SMR performance over a large dataset: N= 80) that we call the SMR predictor. This predictor can be determined from a two-minute recording of a 'relax with eyes open' (EO) condition that is recorded before the user training and instructions. During this condition, the user is not involved in any task. This predictor quantified the potential for

desynchronization of the SMRs during motor imagination. In the same study, they computed the SMR predictor using recordings of a 'relax with eyes closed' (EC) condition (in which the alpha band increase over occipito-parietal areas) as they observed that the SMR predictor can be distorted by contributions from the occipital visual rhythm, which has a peak in the alpha frequency range and is typically much stronger than the SMR rhythm. Their results were substantially worse suggesting a low contribution of occipital alpha. In 2014, Suk et., al [174] aim was to find spatio-temporal patterns in resting-state EEG data using Bayesian Spatio-Spectral Filter Optimization (BSSO) to predict BCI users performance. They found that the higher the classification performance, the larger the values found in the probability density function around the mu- and beta-bands. Those results suggest that there might exist a predictor similar to the SMR-predictor, located in the beta frequency band.

Thereafter, the relationship between other frequency bands and SMR or users' performance have been studied. Grosse-Wentrup et. al [69] reported a relationship between the γ band and the SMRs. SMRs were positively correlated with the power of frontal and occipital γ oscillations, and negatively correlated with the power of centro-parietal γ oscillations. Furthermore, They suggested that the attentional processes, giving rise to the γ rhythms, have a causal influence on SMRs. According to the literature [167–169], γ oscillations are related to the frontal-parietal network, which is itself connected to the mirror neuron system (MNS). The MNS has been demonstrated to be involved in action recognition, MI and imitation of simple movements, which are already present in the motor repertoire of the acting individuals [175, 176]. Therefore, the γ band is related to movement memory during MI. To conclude, the γ band has found to be related to SMRs but also MNS both involved in MI and therefore could be a interesting factor to better understand MI-BCI performances.

Later, Ahn et. al, [177] used the task of relaxing with EO again to investigate the impact of theta waves when doing motor imagery. In this study high θ and low α EEG waves were observed in users with difficulty controlling a BCI. Results showed a positive correlation between offline accuracy and the α-band power and a negative correlation in the θ-band. In addition, they proposed a new predictor, the performance potential (PP) factor, that was designed combining different frequency bands (correlation of r= 0.59, $p < 1.0e^{-}6$ with BCI performance over a large dataset: N= 61). See Section 6.2.3.2 for more details. In 2020, Kwon et. al, [178] suggested that the use of multiple brain states may yield a more robust predictor than the use of a single state alone, i.e., combining data from EO and EC resting states. They proposed a new predictor, the resting state predictor (RSP), that stronlgy correlates with online BCI performance (correlation of r= 0.71, $p < 1.0e^{-}7$ with BCI performance). See Section 6.2.3.3 for more details.

However, all those predictors have limits:

- To compute the SMR predictor, a heavy numerical procedure is needed. First, the computation of a fit of the 1/f noise spectrum is needed and then each power spectral density (PSD) curves are modeled to be decomposed into two components: the noise component and the two peak (α and β) components. The determination of those two components curve-fitting parameters is not trivial.

- The PPfactor, which is based on a simple formula, is based on four spectral bands powers and the formula is based on observations made on one single experiment of one session where 52 subjects participated. However, mental states such as attention, fatigue and emotion among others might have an impact on those spectral powers making it difficult to find a subject-specific reliable PPfactor. In addition, it only uses data from EO condition, which is impacted by the occipital visual rhythm and cognitive activities.

- The RSP predictor followed the approach of the PPfactor but used both EO and EC resting states, i.e., multiple states. However, only 15 participants participated in the study for a total of 41 sessions and, during their analysis, they considered that all sessions where independent (i.e., they did not took into account the dependence induced when using one same subject for multiple sessions). Therefore, results might have been impacted by this decision as within-subjects performance variability is not taken into account. Furthermore, as for the PPfactor,

Table 6.1: Studies concerning neurophysiological predictors of MI-BCI performance.

Study	Number of subjects	Number of online sessions	Task Type	Support	Predictor
Blankertz et. al, 2010	80	1	Left Hand Right Hand	Resting state OE	✓ SMR-predictor (+)
Ahn et. al, 2013	61	1	Left Hand Right Hand	Resting state OE	✓ Potential Performance factor (PP) (+)
Edelman et. al, 2019	22	8	Left Hand Right Hand Left + Right Hand	Target tracking	✓ User engagement (+)
Suk et. al, 2014	83	1 *(+ 3 calibration sessions)*	Left Hand Right Hand Left + Right Hand	Resting state OE	✓ Bayesian Spatio-Spectral Filter Optimization
Saha et. al, 2019	5	4 *(no Feedback)*	Right Hand Right Foot	Online tasks	✓ Cortical regions of interest
Darvishi et. al, 2018	10	4	Left Hand Right Hand	Reaction time test	✓ Reaction time (-)
Kwon et. al, 2020	15	1 to 5 sessions	Left Hand Right Hand Foot	Resting state OE+CE	✓ Resting state predictor (+)

the formula was based on previous analysis made on the same dataset, which raises questions about robustness.

Finally, in addition to predictors based on resting state baselines, i.e., calculated before the user training, some predictors have been found using specific tasks such as target tracking [179] to assess user engagement using EEG. One could therefore consider that this factor is a neurophysiological factor reflecting a cognitive factor.

In this study, our objective is first to use the exposed predictors to see if they generalize across experiments and datasets. Then, we aim at identifying new potential predictors. Finally, to see if multiple predictors together can improve the performance prediction, we created a computational model that could predict the performances of a BCI user based on multiple neurophysiological features from that user.

6.2 Materials and methods

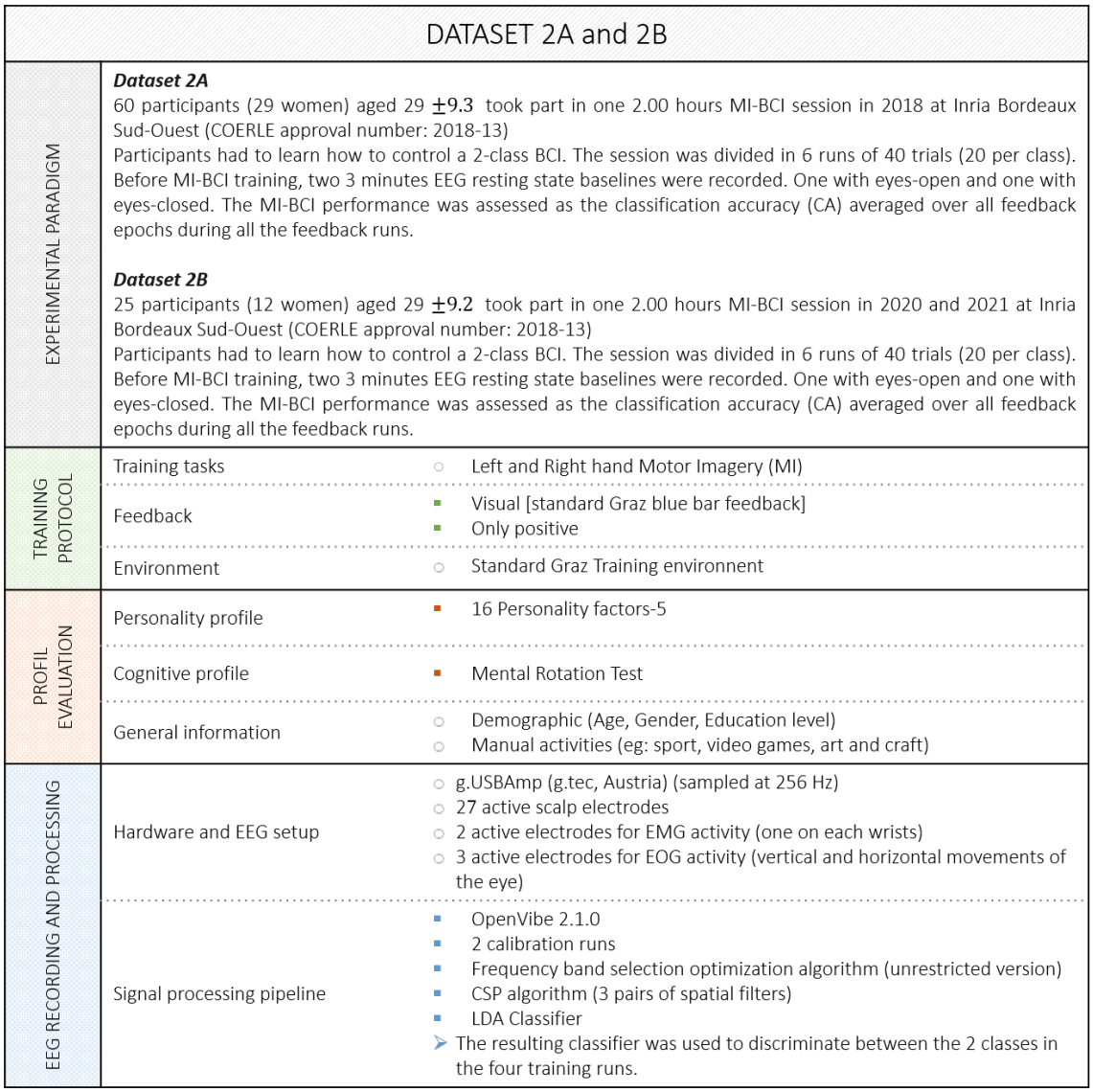
6.2.1 Datasets

For this study we use the Datasets 2A and 2B where a frequency band selection algorithm was used for online experiments, during the calibration before using the CSP algorithm (see Chapter 8 for more details about this algorithm). Figure 9.3 summarizes both datasets.

6.2.2 Raw data signal processing

Before extracting features from our EEG recordings, we corrected or removed biological artifacts from raw data. Such artifacts were head movements, eye blinks or eye movements.

– We automatically corrected eye blinks and eye movements using independent component analysis (ICA) [180] and the Python MNE package [181]. ICA separates a multivariate signal into its additive subcomponents, or sources. Assumptions are that sources are statistically independent and that the values in each source underlie non-Gaussian distributions. [182]. We thus obtain a decomposition into independent components, and the artifact's contribution is localized in only a small number of components. These components (e.g., eye blinks) have to be correctly



DATASET 2A and 2B

EXPERIMENTAL PARADIGM	***Dataset 2A*** 60 participants (29 women) aged 29 ±9.3 took part in one 2.00 hours MI-BCI session in 2018 at Inria Bordeaux Sud-Ouest (COERLE approval number: 2018-13) Participants had to learn how to control a 2-class BCI. The session was divided in 6 runs of 40 trials (20 per class). Before MI-BCI training, two 3 minutes EEG resting state baselines were recorded. One with eyes-open and one with eyes-closed. The MI-BCI performance was assessed as the classification accuracy (CA) averaged over all feedback epochs during all the feedback runs. ***Dataset 2B*** 25 participants (12 women) aged 29 ±9.2 took part in one 2.00 hours MI-BCI session in 2020 and 2021 at Inria Bordeaux Sud-Ouest (COERLE approval number: 2018-13) Participants had to learn how to control a 2-class BCI. The session was divided in 6 runs of 40 trials (20 per class). Before MI-BCI training, two 3 minutes EEG resting state baselines were recorded. One with eyes-open and one with eyes-closed. The MI-BCI performance was assessed as the classification accuracy (CA) averaged over all feedback epochs during all the feedback runs.	
TRAINING PROTOCOL	Training tasks	○ Left and Right hand Motor Imagery (MI)
	Feedback	▪ Visual [standard Graz blue bar feedback] ▪ Only positive
	Environment	○ Standard Graz Training environnent
PROFIL EVALUATION	Personality profile	▪ 16 Personality factors-5
	Cognitive profile	▪ Mental Rotation Test
	General information	○ Demographic (Age, Gender, Education level) ○ Manual activities (eg: sport, video games, art and craft)
EEG RECORDING AND PROCESSING	Hardware and EEG setup	○ g.USBAmp (g.tec, Austria) (sampled at 256 Hz) ○ 27 active scalp electrodes ○ 2 active electrodes for EMG activity (one on each wrists) ○ 3 active electrodes for EOG activity (vertical and horizontal movements of the eye)
	Signal processing pipeline	▪ OpenVibe 2.1.0 ▪ 2 calibration runs ▪ Frequency band selection optimization algorithm (unrestricted version) ▪ CSP algorithm (3 pairs of spatial filters) ▪ LDA Classifier ➢ The resulting classifier was used to discriminate between the 2 classes in the four training runs.

Figure 6.4: Summary table for datasets 2A and 2B.

identified and removed. We used the $ICA.find\_bads\_eog$ function of MNE that detects EOG related components using correlation. We thus obtained a list of sources indices to exclude when re-mixing the data in the ICA.apply() method, i.e. artifactual ICA components (see the Repairing artifacts with ICA tutorial of the MNE API).

– We created 10s epochs with 3s overlap to automatically remove noisy epochs and participants with noisy laplacian channels C3; FC3; CP3; C5; C1; C4; FC4; CP4; C6; C2 as they were used to compute predictors. We used the pyhton Autoreject package [183]. The objective of the algorithm is to select a threshold that will reject noisy epochs: for each potential threshold epochs are split into K folds. Each of the K parts will be a "test" set once, while the remaining $K1$ parts will be combined to be the corresponding "train" set. Then, for each fold K, the threshold is apply to reject trials in the train set. The mean of the signal (for each sensor and timepoint) over the not rejected trials in the train set and the median of the signal (for each sensor and timepoint) over all trials in the test set are computed. The error is calculated between both signals using the Frobenius norm of their difference [184]. For each potential threshold, the mean error of the K folds errors is computed. Finally, the candidate threshold with the lowest error is the best rejection threshold for a global rejection. To remove noisy

Table 6.2: Outlier detection: Group details

Group	number of subjects removed in EO baseline	number of subjects removed in EC baseline
2A	4/60	7/60
2B	0/25	0/25
2A+2B	4/85	7/85

channels, and therefore subjects with noisy laplacian channels, we used the RANSAC from the PREP pipeline [185]. Table 6.2 gather information about removed subjects from our datasets.

6.2.3 Computation of predictors from baseline analyses based on the literature

For all predictors, from the three minutes of baseline recording, we only used the middle two minutes to avoid user movement at the beginning and end of the recording. In addition, we created two Laplacian channels around C3 $[4C3 - FC3 - CP3 - C5 - C1]$ and C4 $[4C4 - FC4 - CP4 - C6 - C2]$ to compute predictors.

6.2.3.1 SMR predictor

We first computed the SMR predictor proposed in Blankertz et., al [186]. We first used the raw EEG data from the EO condition, and used band-pass filtering between 2Hz to 40Hz to reduce physiological noise and power line noise that might affect the EEG recording. From these data, we computed the power spectral density (PSD) between 2Hz to 30 Hz. The PSD were computed for each of the two Laplacian channels. As proposed in [186], we modeled the PSD curve by using the sum of two functions:

$$g(f;\lambda,\mu,\sigma,\mathbf{k}) = g_1(f;\lambda,k) + g_2(f;\mu,\sigma,k) \tag{6.1}$$

$$g_1(f;\lambda,\mathbf{k}) = k_1 + \frac{k_2}{f^{\lambda}} \tag{6.2}$$

$$g_2(f;\mu,\sigma,\mathbf{k}) = k_3\varphi(f;\mu_1,\sigma_1) + k_4\varphi(f;\mu_2,\sigma_2) \tag{6.3}$$

where $\mathbf{k} = (k_1,k_2,k_3,k_4) \in \mathbb{R}^4, \lambda \in \mathbb{R}$ and $\varphi(\,\cdot\,;m,s)$ indicates the probability density function of a normal distribution with mean m and standard deviation s. Function g_1 models the noise spectrum and g_2 models the peaks around the μ (8-15Hz) and β (15-30Hz) bands. Therefore μ_1 and μ_2 represent the μ and β peaks locations, σ_1 and σ_2 their scale and k_1 and k_2 their amplitude. The nine parameters are estimated by minimizing the L_2-norm of the difference vector $PSD(\mathbf{f})$ - $g(\mathbf{f};\lambda,\mu,\sigma,\mathbf{k})$, where $\mathbf{f}$ is the vector of all available frequency values for the PSD. The SMR predictor is defined as $max_{f\in\mathbf{f}}g_2(\mathbf{f};\mu,\sigma,\mathbf{k}) - g_1(\mathbf{f};\lambda,\mathbf{k})$. Here, we first forced $\mathbf{f}$ to be in the range of the μ band frequencies, which means that we search for the maximum difference between g_2 to g_1 only between 5Hz to 15Hz. Then, we forced $\mathbf{f}$ to be in the range of the β band frequencies (between 15Hz to 30Hz). We chose 15Hz instead of 13Hz to avoid having the β-band predictor within the SMR predictor peak.

6.2.3.2 Performance potential (PP) factor

After removing artifacts from the raw data, the resting state EEG data of the C3 and C4 channels from the recordings in a EO condition were band-pass filtered from 2 to 50Hz. The spectral band power of θ (4-8Hz), α (8-13Hz), β (13-30HZ) and low γ (30-50Hz) were obtained and normalized by using the total spectrum power (4-50Hz) to reduce inter-subject power amplitude variability. Then, the following formula was used:

$$PPfactor = \frac{\alpha+\beta}{\theta+\gamma} \tag{6.4}$$

6.2.3.3 Resting state predictor (RSP)

To compute the RSP factor, we proceed in the same way as the computation of the PP factor but we used both the resting state EEG data from the recordings in a EO and EC conditions:

$$RSP = \frac{\alpha_{EO} + \alpha_{EC}}{\theta_{EO} + \beta_{EO} + \beta_{EC}} \tag{6.5}$$

where EC and EO stand for eyes closed and eyes open.

6.2.4 Variable to predict: BCI Performances

The metric used for online performance was the the online trial-wise accuracy, i.e. the default accuracy measure provided online in the MI-BCI scenarios of OpenViBE, which is also the performance metric that the experimenters were seeing online. This metric is computed by first summing the LDA classifier outputs (distance to the separating hyperplane) over all epochs (1s long epochs, with 15/16 s overlap between consecutive windows) during the trial feedback period. If this sum sign matched the required trial label, i.e., negative for left hand MI and positive for right hand MI, then the trial was considered as correctly classified, otherwise it was not. Finally, the online trial-wise accuracy for each run was estimated as the percentage of trials considered as correctly classified using this approach.

6.2.5 Analyses

6.2.5.1 Correlations

We proposed and studied neurophysiological predictors from the literature but also new neurophysiological predictors. We performed regression analyses for each neurophysiological feature, to estimate how well they were correlated to performances. As we performed numerous tests, we performed a correction for multiple testing using the Benjamini-Hochberg (BH) procedure [145] to control the false discovery rate (FDR).

6.2.5.2 Models of prediction and features

In our study, we extracted features from the PSD curve using the SMR predictor method of each subject during the recording of the baseline (both EO and EC conditions). These features describe the PSD properties, and are used as input of the model to predict the performance. All features were normalized using a z-score normalization. As described in 6.5, the features we extracted are the following:

- **Estimated λ parameter** that was computed for function g1, the model for the noise spectrum.
- **SMR predictor** that was computed by:
  $max_{f \in \mathbf{f}} g_2(\mathbf{f}; \mu, \sigma, \mathbf{k}) - g_1(\mathbf{f}; \lambda, \mathbf{k})$, where $\mathbf{f}$ is in the μ-band frequency range (5 to 15 Hz).
- **β-band predictor** computed similarly to the SMR predictor, as $max_{f \in \mathbf{f}} g_2(\mathbf{f}; \mu, \sigma, \mathbf{k}) - g_1(\mathbf{f}; \lambda, \mathbf{k})$, where $\mathbf{f}$ is in the β-band frequency range (15 to 30 Hz).

Furthermore, we also investigated how those features varies across time. These features are new and proposed for the first time in this thesis. Indeed, their variability might explain differences in mean peak amplitude or could be related to BCI performances:

- **Variance of the Estimated λ parameter across epochs**
- **Variance of the SMR predictor across epochs**
- **Variance of the β-band predictor across epochs**

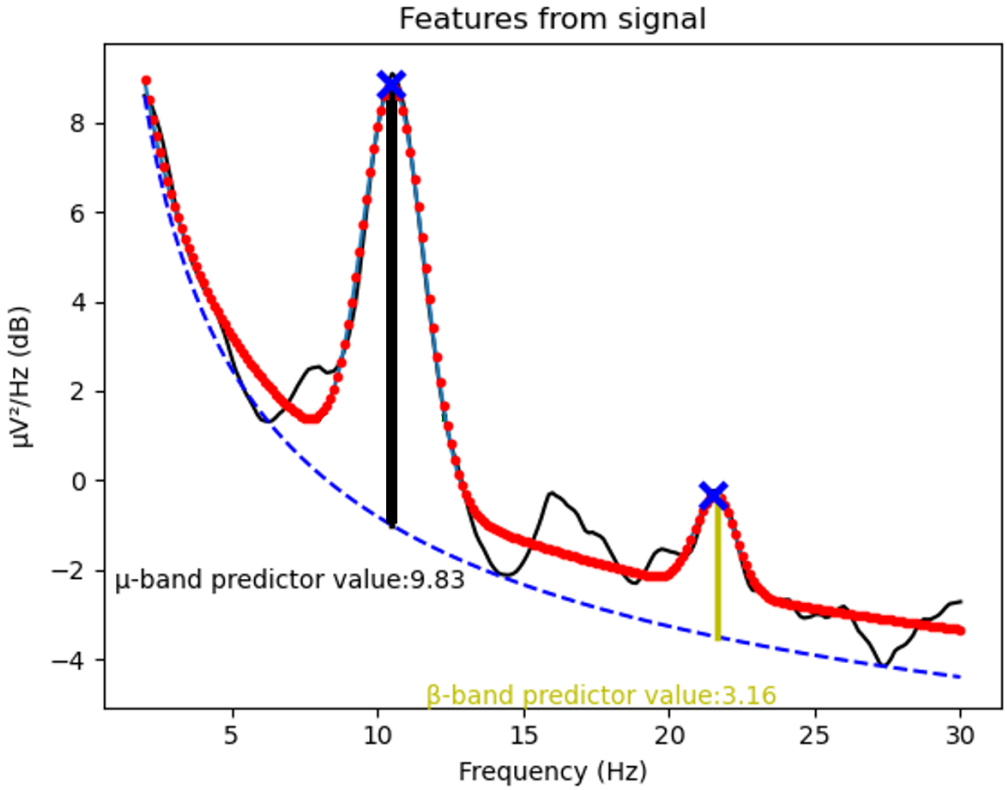


Figure 6.5: In the plot, we can see most of the features that were extracted from the signal. The solid black signal is the original power spectral density of the user, the dashed blue signal is the estimated floor noise ($g_1(f;\lambda,\mathbf{k})$) while the dotted red signal is the fitted values $g(f;\lambda,\mu,\sigma,\mathbf{k})$. The two horizontal lines describe the predictors, the black one is for SMR predictor and the yellow one is for the β-band predictor. The blue 'X' represents the frequency of the peak.

In addition to those factors, we added powers of different frequency bands: θ (4-8Hz), α (8-13Hz), β (13-30HZ) and low γ (30-50Hz).

In order to build a statistical model to predict/explain MI-BCI performances we used a LASSO regression to obtain models that could predict the performances of MT-BCI users from the neuro-physiological characteristics of the user. See Table 6.3 (see Chapter 4 for more details about LASSO regression and methodology). In total we used 20 features from both EO and EC baselines.

Table 6.3: Summary table explaining the statistical model used to predict/explain MI-BCI performances.

STATISTICAL MODEL OF PREDICTION	
METHOD	LASSO
EXPLANATORY FEATURES	▪ Estimated λ parameter (EO and EC) ▪ SMR predictor (EO and EC) ▪ β-band predictor (EO and EC) ▪ Variance of the estimated λ parameter (EO and EC) across epochs ▪ Variance of the SMR predictor (EO and EC) across epochs ▪ Variance of the β-band predictor (EO and EC) across epochs ▪ Theta (θ) power (EO and EC) ▪ Alpha (α) power (EO and EC) ▪ Beta (β) power (EO and EC) ▪ Gamma (γ) power (EO and EC)
VARIABLE TO PREDICT	The online trial-wise accuracy, i.e. the default accuracy measure provided online in the MI-BCI scenarios of OpenViBE
EVALUATION OF STABILITY AND RELIABILITY	▪ Leave-one-out cross validations ▪ Random model with 1000 permutations

6.3 Result

6.3.1 Outliers detection

We excluded from the analyses the subjects whose mean classification accuracy was above or below two standard deviations (SD) from the group's mean performance (see summary in Table 6.4).

Table 6.4: Outlier detection: Group details. Subjects with corrupted channels among the laplacians' were removed prior this analysis.

Group	$Perf_{Group}$	SD	Number of Outliers	Final number of subjects (Between-subject analyses)
2A	63.41%	16.47 %	3	50/53
2B	64.90%	15.13 %	1	24/25
2A+2B	63.78%	15.81 %	4	74/78

6.3.2 Correlations

We conducted regression analysis between the PPfactor, the RSP, the SMR predictor and online performance (see Figure 6.6) for each group (i.e., Dataset 2A, Dataset 2B and Dataset 2A+2B). After adjustment of the pvalues using BH procedure, we found that all predictors were strongly positively correlated to online performances (see Table 6.5) for all groups. As correlations between PPfactor and the RSP and online performances were stronger using Laplacian channels C3 and C4, we used those values for the rest of our analyses.

In addition, we observed that all predictors were strongly correlated to each other (see Figure 6.7).

This result, notably the correlation between the RSP and the PPfactor, suggested that β and γ bands impact is negligible compared to α and θ bands (see Equations 6.4 and 6.5). This result could be expained by the fact that the EEG spectrum follow an 1/f curve. Therefore, we decided to perform additional correlation analyses using the α/θ ratio for both EO and EC baselines (see Figure 6.8) which yielded a significant positive correlation with online performances of r_{EO}=0.58, r_{EC}=0.58 ($p_{adj} < 10^-4$, $p_{adj} < 10^-4$), r_{EO}=0.66, r_{EC}=0.17 ($p_{adj} < 10^-2$, $p_{adj} = 0.4$) and r_{EO}= 0.60, r_{EC}=0.47 ($p_{adj} < 10^-5$, $p_{adj} < 10^-4$) for Dataset A, B and A+B respectively.

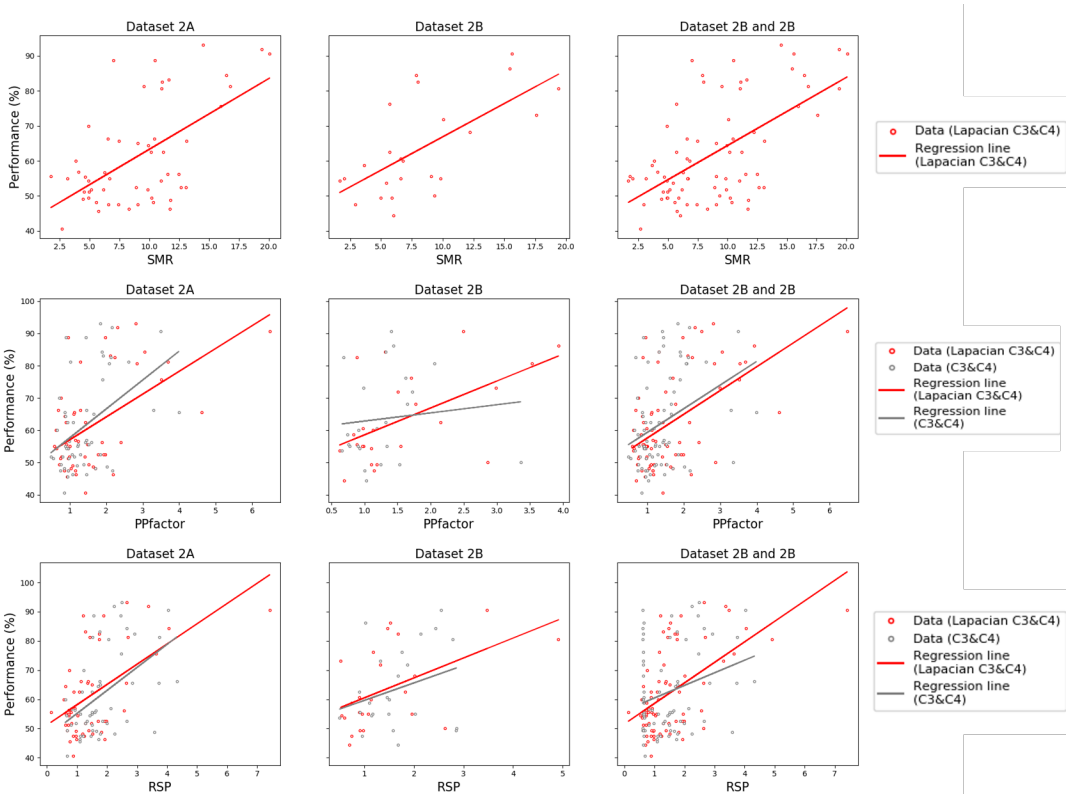


Figure 6.6: Regression analysis between neurophysiological predictors of the literature (PPfactor, RSP, SMR) and online BCI classification accuracy using Dataset 2A, Dataset 2B and Dataset 2A+2B

Factor / Group	SMR		PPFactor		RSP	
	Laplacian C3 and C4	C3 and C4	Laplacian C3 and C4	C3 and C4	Laplacian C3 and C4	C3 and C4
2A	R= 0.58 $p_{adj}<10^{-4}$	-	R=0.53 $p_{adj}<10^{-4}$	R=0.45 $p_{adj}<10^{-2}$	R=0.56 $p_{adj}<10^{-4}$	R=0.48 $p_{adj}<10^{-3}$
2B	R= 0.66 $p_{adj}<10^{-3}$	-	R=0.56 $p_{adj}<10^{-2}$	R=0.11 $P_{adj}=0.6$	R=0.50 $p_{adj}<0.05$	R=0.29 $P_{adj}=0.18$
2A + 2B	R= 0.60 $p_{adj}<10^{-6}$	-	R=0.54 $p_{adj}<10^{-5}$	R=0.35 $p_{adj}<10^{-2}$	R=0.56 $p_{adj}<10^{-5}$	R=0.28 $p_{adj}<0.05$

Table 6.5: Correlation results between each factor (calculated using both C3 and C4 and laplacian of C3 and C4) and online performance for each group. pvalues are adjusted using BH procedure. We did not computed the SMR predictor using only C3 and C4 as the predictor was initially proposed using Laplacians [**?**]

6.3.3 Models of prediction

The results indicated that the predictions made for Dataset 2A were better than chance ($p = 0.001$), with a mean absolute error $\sum_{i=1}^{N} \frac{|V_{pred(i)} - V_{real(i)}|}{N}$, n being the total number of models generated, of 10.43%. The mean absolute error (in %) of the random models were 11.23% and 11.92% for $p = 0.01$ and $p = 0.05$ respectively (see Chapter 4 for more details about the used method). More precisely, 100% of the generated models included the same six factors: the SMR predictor, the estimated λ (i.e., the 1/f parameter of the noise spectrum) , the α band power from the EO condition baseline and the θ power, the β power and the variance of the SMR predictor across epochs from the EC condition baseline. More than 80% of the generated models included the variance of the SMR predictor across epochs and the γ power from the EO condition baseline and the α band power from the EC condition baseline (see Figure 6.9). Finally, results showed a positive significant correlation r=0.54 ($p < 10^{-4}$) between online performances and predictive performances.

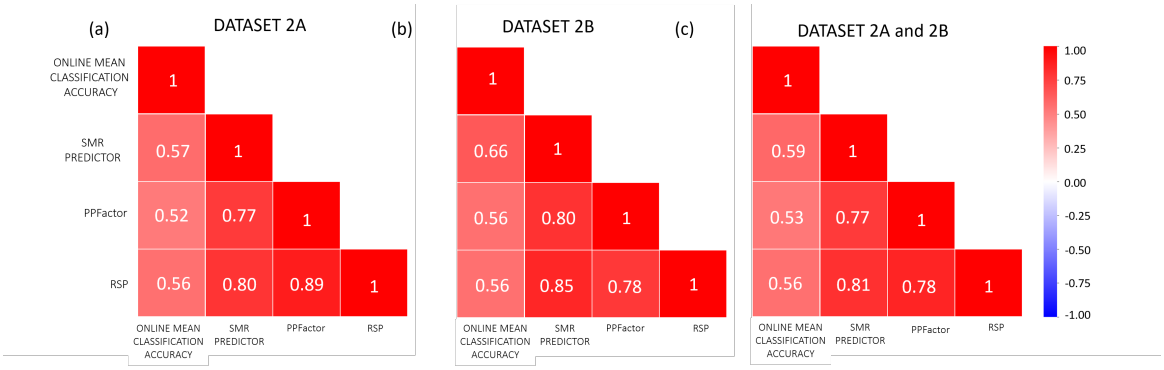


Figure 6.7: Correlation matrices showing correlation score after adjustment of the pvalues (using BH procedure) of Datasets 2A(a), 2B(b), 2A+2B(c). All correlations were significant ($p < 0.05$, at least).

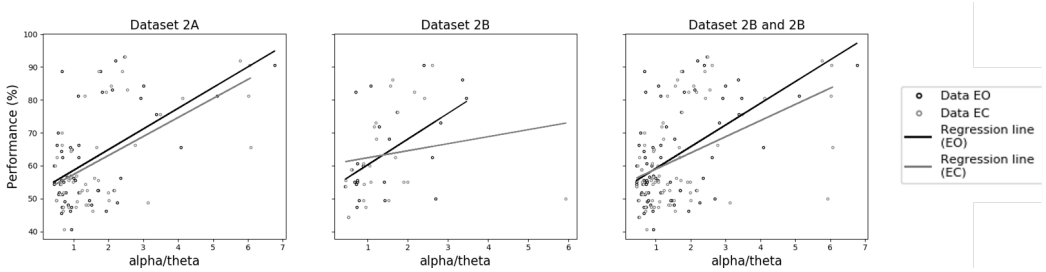


Figure 6.8: Regression analysis between the α, θ ratio (using Laplacians C3 and C4) and online BCI classification accuracy using Dataset 2A, Dataset 2B and Dataset 2A+2B.

Finally, the final model based on all subjects of Dataset 2A is defined as follow (using normalized factors):

$$Performance(\%) = 61.92 + 1.78 \times SMR_{EO} + 2.14 \times \lambda_{EO} + 0.21 \times var(\lambda_{EO}) + 3.15 \times \alpha_{EO} - 0.72 \times \gamma_{EO} - 4.11 \times \theta_{EC} + 1.39 \times \alpha_{EC} - 0.78 \times \beta_{EC} - 1.76 \times var(SMR_{EC}) \quad (6.6)$$

with $var(x)$ the variance of the x factor across epochs.

We computed the predictive performance for each subject of Dataset 2B using Equation 6.6. The mean absolute error $|Perf_{pred} - Perf_{real}|$ was 9.3% (see Figure 6.10) and better than chance ($p = 0.005$) as The mean absolute error (in %) of the random models were 9.64% and 10.61% for $p = .01$ and $p = .05$ respectively. Results showed a positive significant correlation r=0.51 ($p = 0.01$) between online performances and predictive performances.

In order to study each factor individually, we computed a correlation matrix (Figure 6.11) using Dataset 2A+2B (see Appendix for correlation matrices of Dataset 2A and 2B individually).

We can observe that not all factors included in the model are directly correlated to performances. For example, the estimated λ factor is not correlated to performances or any other features but is included in the model with a strong positive weight. Furthermore, the θ of the EC resting state, which had a strong negative impact on our model, only have a coefficient correlation of -0.38 with BCI performances.

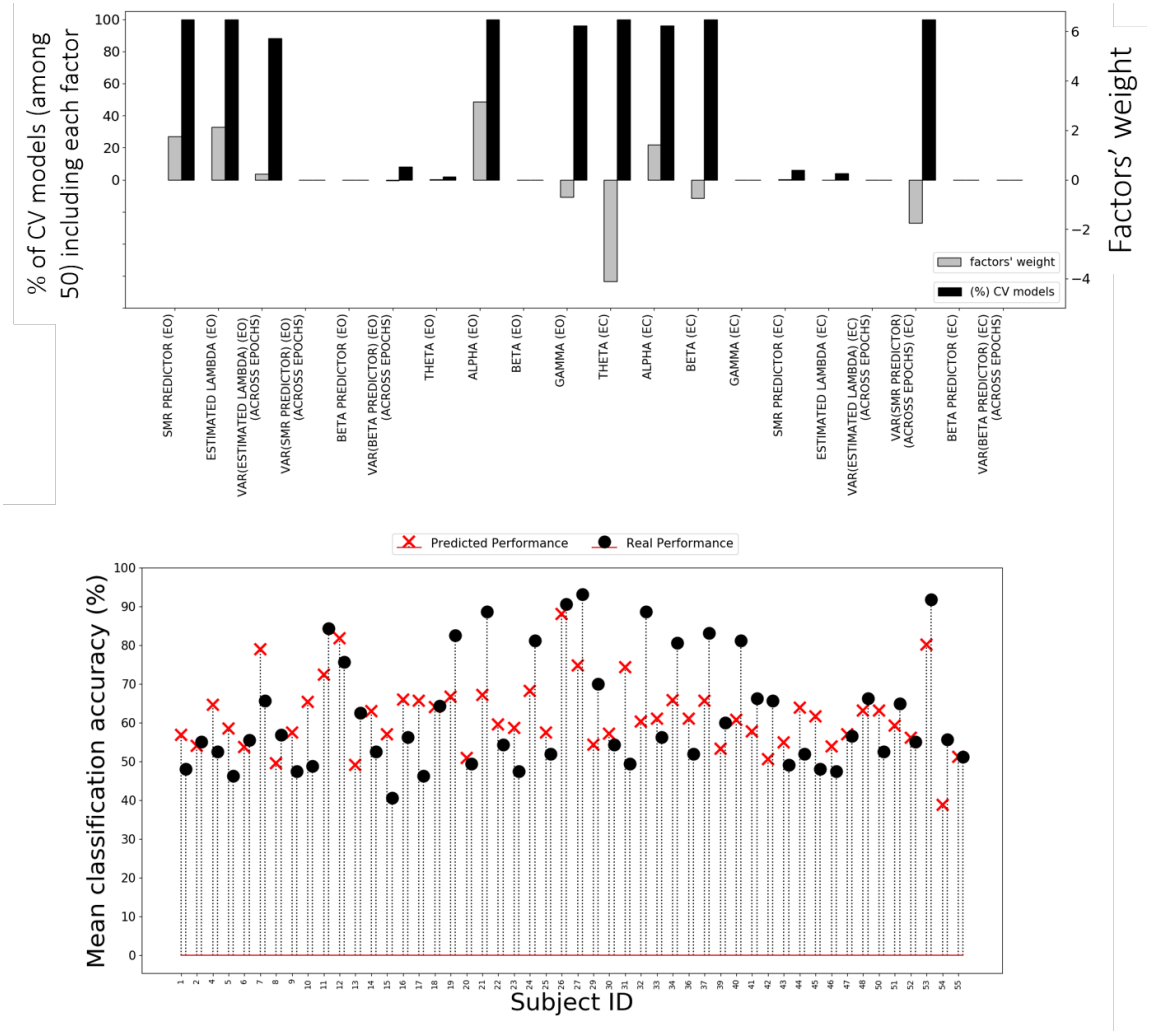


Figure 6.9: Results of the different models generated for Dataset 2A. On the top, the percentage of Cross-Validation models including each factor (black) and their weight (grey). On the bottom, in black (circle), the real performance of each subject and in red (cross), the predicted performance of each subject generated using the model generated from the training dataset (All subjects except the target one).

6.4 Discussion

SMR predictor, PPfactor and RSP

In this work we first studied the reproducibility of previous works that introduced neurophysiological predictor of MI-BCI performance: The SMR predictor [**?**], the PPfactor [177] and the RSP [178]. The SMR predictor was designed with mainly μ-band (10-14 Hz) information. The PPfactor was designed including more frequency bands (θ (4-8Hz), α (8-13Hz), β (13-30HZ) and low γ (30-50Hz)). Both were computed from a single resting state (eyes-open). The RSP was proposed as a combination of EO and EC predictors. All were positively correlated to online performances (the SMR predictor obtaining the highest score) when using Laplacian channels around C3 [$4C3 - FC3 - CP3 - C5 - C1$] and C4 [$4C4 - FC4 - CP4 - C6 - C2$]. In the related studies, the PPfactor and the RSP were computed using C3 and C4 (or Cz) channels only. The use of Laplacian filters [187] allowed us to reduce the background activity by subtracting the average of the surrounding channels from the channel of interest. It is particularly useful to isolate the relevant spatial information embedded in the signals. Therefore, when studying potential predictors of MI-BCI performances, we believe that results would be more accurate using Laplacian filters. This hypothesis has been confirmed by our results as correlations were stronger using Laplacian channels around C3 and C4. Furthermore, all

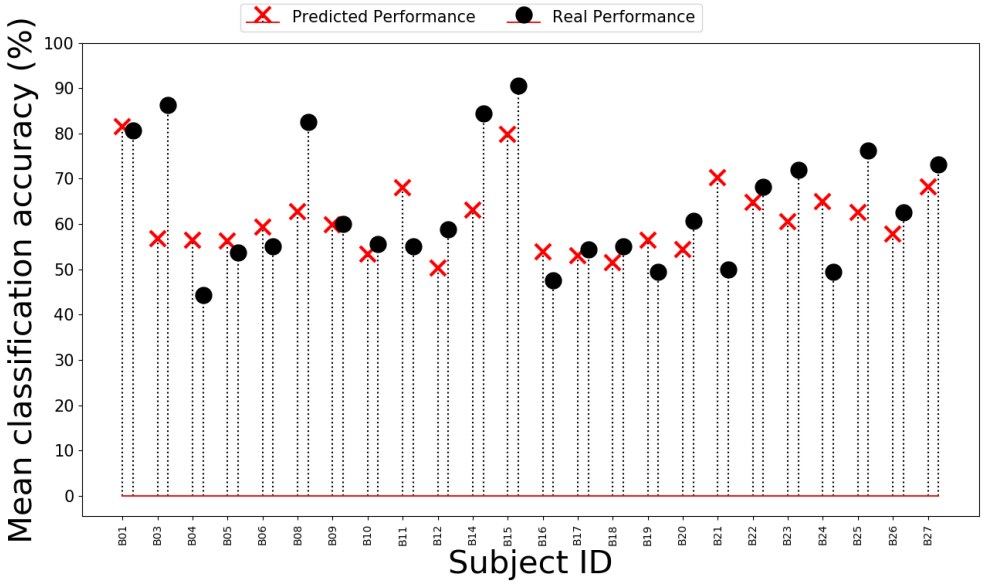


Figure 6.10: In black (circle), the online performance of each subject of Dataset 2B and in red (cross), the predicted performance of each subject generated using the model generated using Equation 6.6.

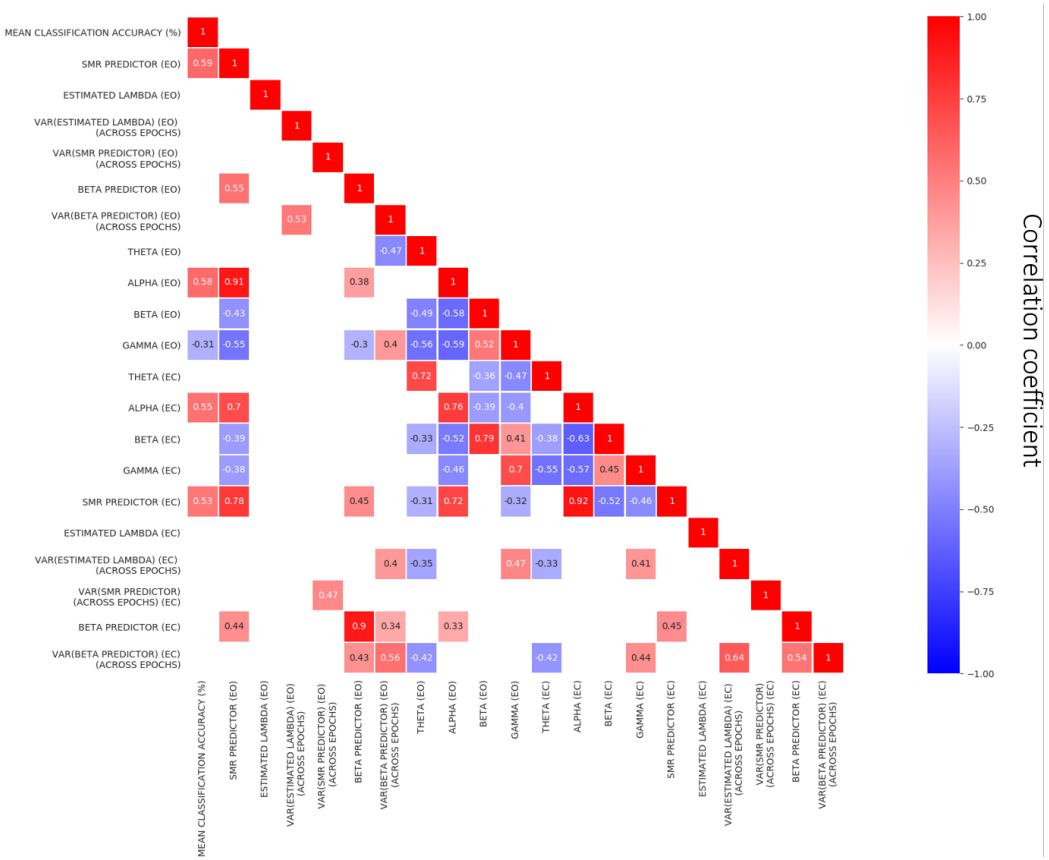


Figure 6.11: Correlation matrix of neurophysiological features and MI-BCI performance extracted from Dataset 2A+2B. Correlation coefficients are reported only for correlations with $p_{adj} < 0.01$.

predictors were positively correlated to each other which means that they probably measure similar effects underlying neurophysiological effects.

Model of prediction

Another objective of this work was to study the feasibility of predicting/explaining MT-BCI performances across experiments, using a statistical model based on the subjects' neurophysiological characteristics only. The objective was, through those models of prediction, to identify the stable factors across experiments that may be related/associated to subject performances and thus, on the longer-term, improve BCI user training protocols. In order to have interpretable models we used the LASSO that promotes sparse solutions. We were able to find a model based on one dataset, significantly better than chance and we could prove its reliability by using it on a new set of data of another experiment which had the same training protocol. Factors with the strongest impact in our model were related to the alpha bands and theta bands using both eyes-closed and eyes-open baselines. More specifically, the SMR predictor, the estimated λ (i.e., the 1/f parameter of the noise spectrum) and the α-band power, all computed using the eyes-open baseline. In addition, factors from the eyes-closed baseline were also selected with a strong weight compared to other factors such as the θ- and α-band powers and the variance of the SMR predictor across time.

This result first enhances the reliability of the SMR predictor as a strong indicator of MI-BCI performances. In addition, the exponent parameter λ of the $1/f^{\lambda}$ noise spectrum (i.e., its slope) was also selected with a strong positive weight in our model. The 1/f noise takes its name from the shape of the spectrum since the PSD is inversely proportional to frequencies. Higher λ (sharper slope) indicates an increase of signal to noise ratio and therefore less neural noisy spiking activity [188] compared to lower λ. Therefore, it is easier to separate neurophysiological mechanism from spiking activity. Increase of spiking activity can be related to aging [189] or attention [190] or cognitive speed [191]. Our results also suggest that an increase of signal to noise ratio at rest might impact positively MI-BCI performances which goes in the direction of previous work [192] that reported that participant that showed difficulties controlling a BCI had a higher noise than those that could control a BCI.

Furthermore, α-band power during eyes-open resting state had a non negligible positive weight in our model as well as the α-band power during eyes-closed resting state which is consistent with the reasoning used to build the SMR predictor [14]. Indeed, during a motor task, the power decrease (ERD) in α-band is interpreted as a reflection of motor cortex activation [173]. Therefore, the higher the difference between the α-band power at rest and the one during an MI task , the more likely it is to successfully control a MI-BCI. Our results also reflected a strong negative impact of the θ-band power at rest with eyes-closed on MI-BCI performances. The production of theta with eyes-closed is a well known accompaniment of states of deep relaxation such as stage 1 sleep, meditation and hypnosis [160]. This result is surprising as studies have reported that, an alpha/theta neurofeedback training with eyes-closed, with the aim of maximising the theta to alpha ratio, could improve creative performance [193, 194]. In addition, subjects who meditate have an higher SMR predictor [195] than those who do not and training in mindfulness meditation improves ability to control a BCI device [196]. One explanation could be, as predicted λ and the θ-band power at rest with eyes-closed are negatively correlated, that higher θ-band power represent a higher noise in the signal. This suggestion combined with our previous result on the 1/f could support the hypothesis that the functional relevance of the 1/f power in resting state neural activities substantiate the necessity of isolating the 1/f component from oscillatory activities when studying the functional relevance of spontaneous brain activities.

In addition, one feature of variability was selected in our model with an important negative weight: the variance of the SMR predictor across epochs in eyes-closed condition. Correlation analyses revealed that this factor was not correlated with any other available factor nor MI-BCI performances. This suggests that, subjects with a more important variability of the *alpha* activity during eyes-closed resting state might have more difficulties to control a BCI system than those with a lower variability. As *alpha* activity can be considered as a marker of internally-directed attention processing [197] this result could reflect that fluctuation of focused attention of a user could impact negatively his performances.

Interestingly enough no factor related to the β-band was selected in our model, which might suggest that its impact on MI-BCI performances is negligible compared to other frequency bands in the motor

cortex.

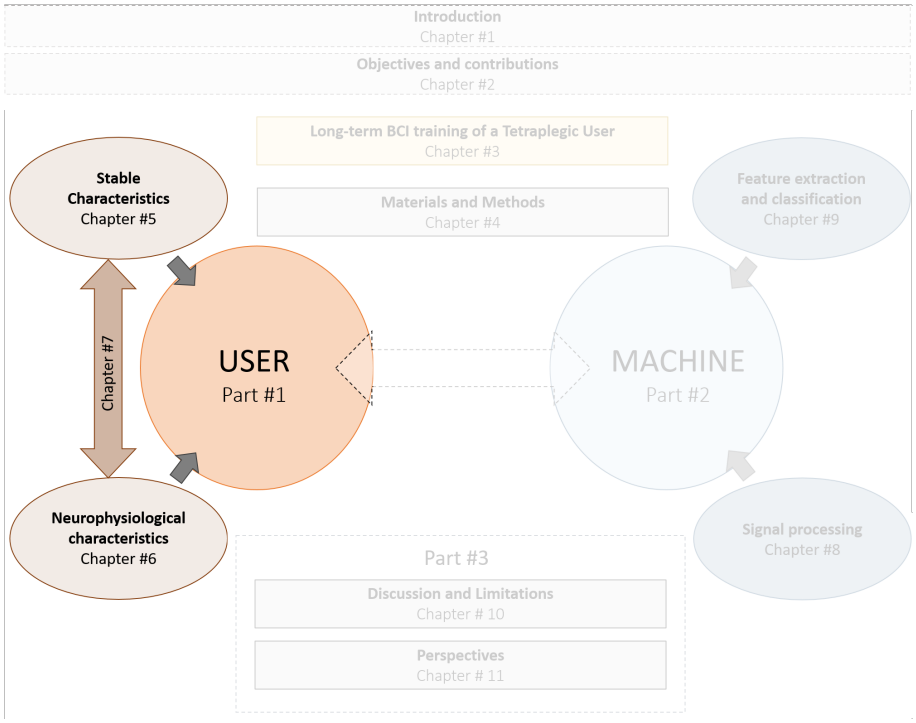
Finally, what is interesting in these results is that a neurofeedback training before an MI-BCI training might impact users' performances. The SMR neurofeedback training aims at increasing the amplitude of the SMRs. However, it seems, according to our results, when using our datasets, that an alpha/theta neurofeedback training eyes-closed could impact negatively MI-BCI performances. Indeed, this training aims to increase the alpha/theta ratio while our results suggest that it should be reduced.

All together, those results suggest that it was possible to predict MI-BCI performances of unseen users using neurophysiological characteristics of a user only. This raises the question of whether combining all user-related factors (i.e., traits and neurophysiological characteristics) can lead to a better model of MI-BCI predictions. In the next Chapter, we will try to answer this question.

7

Combining all users' characteristics

ROADMAP -



QUICK SUMMARY -

The studies conducted in the two previous Chapters suggested that 1) using user traits only could not reliably predict MT-BCI performances and 2) it was possible to predict MI-BCI performances of unseen users using neurophysiological characteristics of a user only. Therefore, the objective of this Chapter was to create computational models of BCI user training that could predict the performances of various BCI users, based on both their traits (i.e., stable characteristics) and their neurophysiological characteristics. To our knowledge, no study has tried to combined those characteristics together in order to build a predictive model of performances. Our results suggested that it was possible to predict MI-BCI performances of unseen users using stable and neurophysiological characteristics of users with an error of 9.61% ($p = 0.037$).

Contents

7.1 Introduction

At the end of Chapter 5, we hypothesised that combining users' traits together with other factors such as neurophysiological characteristics could enable us to create a more robust and precise model of prediction for MT-BCI performances. In addition, we where able to find a stable and reliable model of performance using neurophysiological characteristics in Chapter 6 but not using users' traits only (Chapter 5). Therefore, we found it relevant to combine in a single model all user-related characteristics that we have used previously. If this model is stable and reliable, it could give information about possible relationship between factors and therefore provide additional information on the user's learning.

7.2 Materials and Methods

7.2.1 Datasets

For this study we will use Datasets 2A and 2B where a frequency band selection algorithm was used for online experiments, during the calibration before using the CSP algorithm (see Chapter 8 for more details). Figure 7.1 summarizes both datasets. We chose those two datasets as they were the only one where all users' characteristics were available. One subject from Dataset 2B was removed as he did not completed the 16PF5 personality questionnaire. In addition, seven subjects from Dataset 2A were removed due to corrupted EEG signals during baseline recordings (see Chapter 6 , Section 6.2.2). In total, we used 53 subjects from Dataset 2A and 24 from Dataset 2B.

7.2.2 Analyses

7.2.2.1 Models of prediction and features

In order to build a statistical model to predict/explain MI-BCI performances we used a LASSO regression to obtain models that could predict the performances of MI-BCI users from the stable characteristics (i.e., traits) and neurophysiological characteristics of the user. See Table 7.1.(see Chapter 4 for more details about LASSO regression and methodology). Factors related to the traits of the user are the same ones used in Chapter 5, Section 5.4.1 and factors related to the neurophysiological characteristics of the user are the same as the ones described in Chapter 6, Section 6.2.5.2.

The metric used for online performance was the the online trial-wise accuracy, i.e. the default accuracy measure provided online in the MI-BCI scenarios of OpenViBE, which is also the performance metric that the experimenters were seeing online. Details about this metric are exposed in Chapter 6, Section 6.2.4.

7.2.2.2 Correlations

In addition to the model, we performed correlation analyses for each neurophysiological and stable characteristic (traits) features, to estimate how well they were correlated to performances and to each other using both Datasets 2A and 2B to maximize the number of available data. As we performed numerous tests, we performed a correction for multiple testing using the BH procedure [145] to control the FDR.

DATASET 2A and 2B		
EXPERIMENTAL PARADIGM	***Dataset 2A*** 60 participants (29 women) aged 29 ±9.3 took part in one 2.00 hours MI-BCI session in 2018 at Inria Bordeaux Sud-Ouest (COERLE approval number: 2018-13) Participants had to learn how to control a 2-class BCI. The session was divided in 6 runs of 40 trials (20 per class). Before MI-BCI training, two 3 minutes EEG resting state baselines were recorded. One with eyes-open and one with eyes-closed. The MI-BCI performance was assessed as the classification accuracy (CA) averaged over all feedback epochs during all the feedback runs. ***Dataset 2B*** 25 participants (12 women) aged 29 ±9.2 took part in one 2.00 hours MI-BCI session in 2020 and 2021 at Inria Bordeaux Sud-Ouest (COERLE approval number: 2018-13) Participants had to learn how to control a 2-class BCI. The session was divided in 6 runs of 40 trials (20 per class). Before MI-BCI training, two 3 minutes EEG resting state baselines were recorded. One with eyes-open and one with eyes-closed. The MI-BCI performance was assessed as the classification accuracy (CA) averaged over all feedback epochs during all the feedback runs.	
TRAINING PROTOCOL	Training tasks	○ Left and Right hand Motor Imagery (MI)
	Feedback	▪ Visual [standard Graz blue bar feedback] ▪ Only positive
	Environment	○ Standard Graz Training environnent
PROFIL EVALUATION	Personality profile	▪ 16 Personality factors-5
	Cognitive profile	▪ Mental Rotation Test
	General information	○ Demographic (Age, Gender, Education level) ○ Manual activities (eg: sport, video games, art and craft)
EEG RECORDING AND PROCESSING	Hardware and EEG setup	○ g.USBAmp (g.tec, Austria) (sampled at 256 Hz) ○ 27 active scalp electrodes ○ 2 active electrodes for EMG activity (one on each wrists) ○ 3 active electrodes for EOG activity (vertical and horizontal movements of the eye)
	Signal processing pipeline	▪ OpenVibe 2.1.0 ▪ 2 calibration runs ▪ Frequency band selection optimization algorithm (unrestricted version) ▪ CSP algorithm (3 pairs of spatial filters) ▪ LDA Classifier ➢ The resulting classifier was used to discriminate between the 2 classes in the four training runs.

Figure 7.1: Summary table for Datasets 2A and 2B.

7.3 Results

7.3.1 Outliers detection

We excluded from the analyses the subjects whose mean classification accuracy was above or below two standard deviations (SD) from the group's mean performance (see summary in Table 7.2).

7.3.2 Correlations

Only correlations using Dataset A+B are shown in this section (see Figure 7.2). However, additional correlations analyses (on Datasets A and B) were made and are revealed in Appendix IV. We can note that only neurophysiological predictors i.e., the SMR predictor (EO and EC), the α-band powers (EO and EC) and the γ-band power (EO) are directly correlated to MI-BCI performances. In addition, results revealed interesting correlations between neurophysiological characteristics and user traits. Indeed, the apprehension factor was correlated with the SMR predictor EO (R=0.33, p_{adj}=0.004) and EC (R=0.32, p_{adj}=0.006), and the α-band power EO (R=0.33, p_{adj}=0.008) and EC (R=0.33, p_{adj}=0.007). In addition, the β-band power was correlated to the Age (R=0.30, p_{adj}=0.009

STATISTICAL MODEL OF PREDICTION	
METHOD	LASSO
EXPLANATORY FEATURES	▪ Estimated λ parameter (EO and EC) ▪ SMR predictor (EO and EC) ▪ β-band predictor (EO and EC) ▪ Variance of the estimated λ parameter (EO and EC) across epochs ▪ Variance of the SMR predictor (EO and EC) across epochs ▪ Variance of the β-band predictor (EO and EC) across epochs ▪ Theta (θ) power (EO and EC) ▪ Alpha (α) power (EO and EC) ▪ Beta (β) power (EO and EC) ▪ Gamma (γ) power (EO and EC) ▪ 16pf5 factors ▪ Mental Rotation Score ▪ Age ▪ Gender
VARIABLE TO PREDICT	The online trial-wise accuracy, i.e. the default accuracy measure provided online in the MI-BCI scenarios of OpenViBE
EVALUATION OF STABILITY AND RELIABILITY	▪ Leave-one-out cross validations ▪ Random model with 1000 permutations

Table 7.1: Summary table explaining the statistical model used to predict/explain MI-BCI performances.

Table 7.2: Outlier detection: Group details. Subjects with corrupted channels among the laplacians' were removed prior this analysis.

Group	$Perf_{Group}$	SD	Number of Outliers	Final number of subjects (Between-subject analyses)
2A	63.41%	16.47 %	3	50/53
2B	63.54%	13.82 %	1	23/24
2A+2B	62.08%	14.27 %	4	73/77

(EO), R=0.41, p_{adj}=0.0003 (EC)) but also the vigilance factor (R=-0.33, p_{adj}=0.004 (EO), R=0.30, p_{adj}=0.0009 (EC)).

7.3.3 Models of prediction

The results indicated that the predictions made for Dataset 2A were better than chance ($p \leq 0.05$), with a mean absolute error of 10.55%. The mean absolute error (in %) of the random models were 10.54% and 11.77% for $p = .01$ and $p = .05$ respectively (see Figure 7.3). In this model, several factors representing the user's traits were selected in, at least, 80% of the generated models i.e., the mental rotation score (with a negative weight), the reasoning factor (with a positive weight), the dominance factor (with a negative weight), the sensitivity factor (with a positive weight), the openness to change factor (with a negative weight), the perfectionism factor (with a negative weight) and the tension factor (with a negative weight). In addition, models also included users' neurophysiological factors i.e., the estimated λ factor (EO), the α-band power (EO and EC), the γ-band power (EO), the θ-band power (EC) and the variance of the SMR predictor (EC) across epochs. Finally, results showed a positive significant correlation r=0.50 ($p = 2.10^{-4}$) between online performances and predicted performances.

Finally, the final model based on all subjects of Dataset 2A was defined using 15 normalized factors:

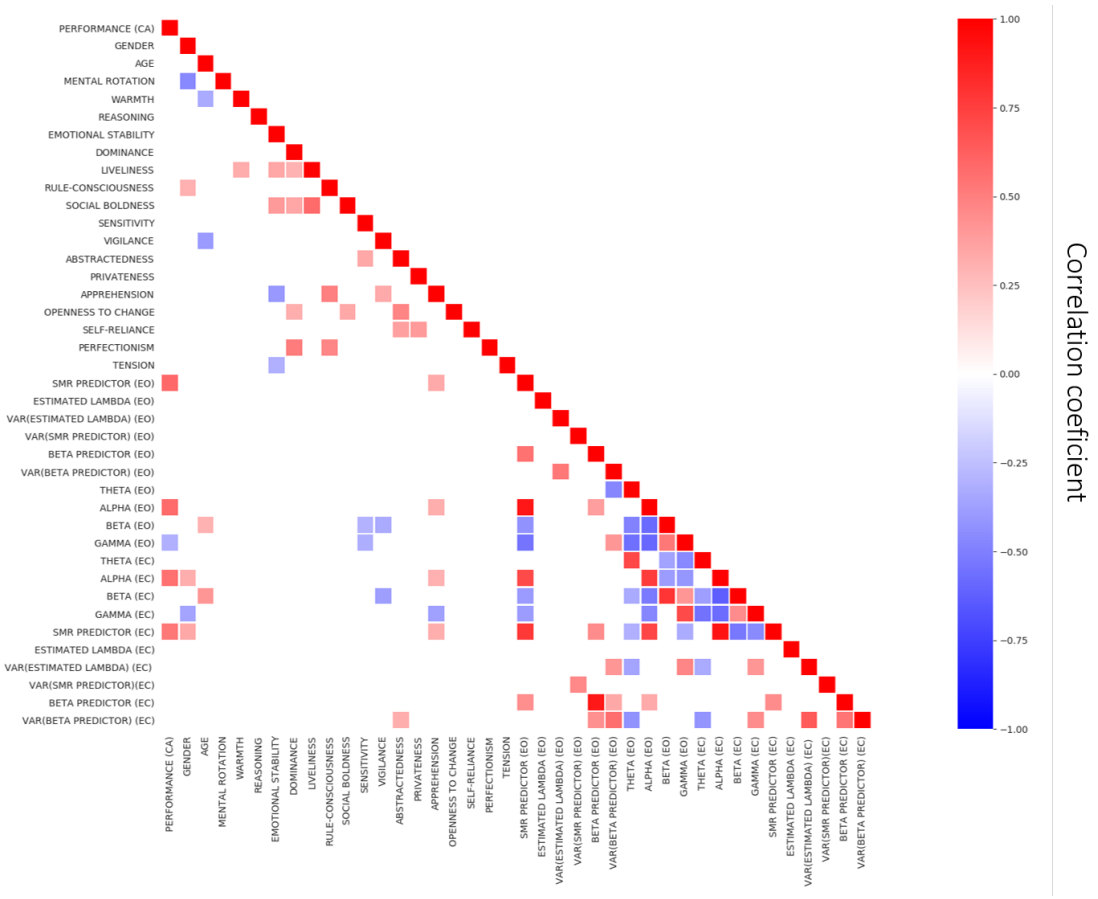


Figure 7.2: Correlation matrix of all user-related features and MI-BCI performance extracted from Dataset 2A+2B. Correlation coefficients are reported only for correlations with $p_{adj} < 0.01$.

$$\begin{aligned} Performance(\%) = 61.92 &- 1.25 \times \text{MENTAL ROTATION} + 1.87 \times \text{REASONNING} - 0.38 \times \text{DOMINANCE} \\ &+ 0.18 \times \text{SENSITIVITY} - 1.52 \times \text{OPENESS TO CHANGE} - 0.47 \times \text{PERFECTIONISM} \\ &- 2.26 \times \text{TENSION} + 2.08 \times \lambda_{EO} + 4.01 \times \alpha_{EO} - 0.32 \times \gamma_{EO} \\ &- 3.77 \times \theta_{EC} + 2.31 \times \alpha_{EC} + 0.02 \times SMR_{EC} - 1.42 \times var(SMR_{EC}). \end{aligned} \quad (7.1)$$

Using the model built on Dataset 2A, we computed the predictive performance for each subject of Dataset 2B using Equation 7.1. The mean absolute error $|Perf_{pred} - Perf_{real}|$ was 9.61% (see Figure 7.4) and better than chance ($p = 0.037$) as the mean absolute error (in %) of the random models were 8.38% and 10.05% for $p = .01$ and $p = .05$ respectively. In addition, results showed a positive significant correlation r=0.43 ($p = 0.04$) between online performances and predicted performances.

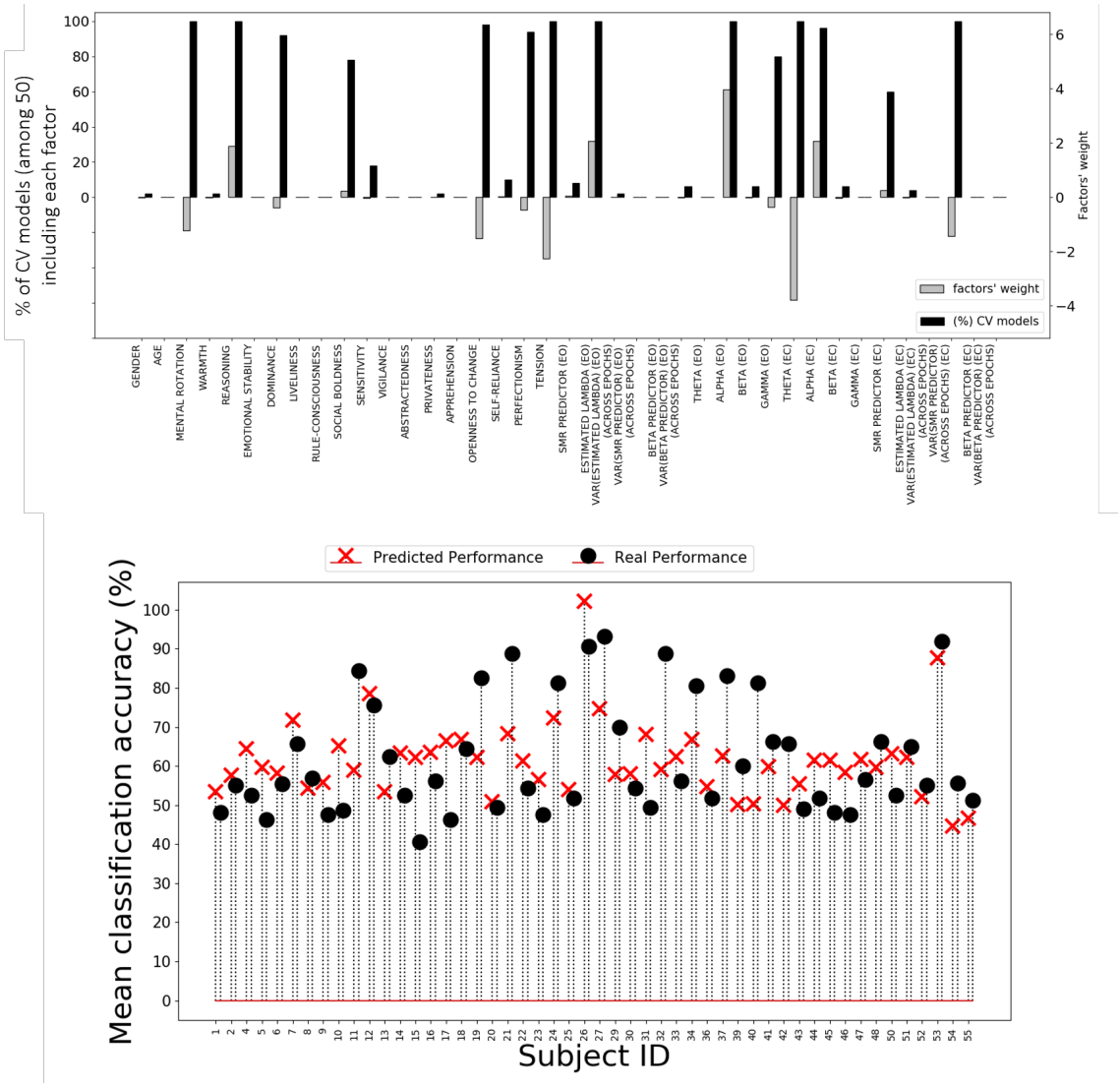


Figure 7.3: Results of the different within-subjects models generated for from Dataset 2A using neurophysiological and stable characteristics of the user. On top, the percentage of leave-one-out cross-validation models including each factor (black) and their weight (grey). Bellow, in black (circle), the real performance of each subject and in red (cross), the predicted performance of each subject generated using the model defined from the training dataset (All subjects except the target one). Finally, in the right plots, the correlation between the real and predicted performances.

7.4 Discussion

When comparing these results to the ones obtained in Chapter 6, Section 6.3.3, we can note that adding stable characteristics to the model (i.e., traits of the user), did not improve the mean absolute error of the model. Indeed the error of the neurophysiological model (see Equation 6.6) was 10.43% and the error of the model combining all factors (i.e., described in this Chapter, see Equation 7.1) was 10.54%. Furthermore, using only neurophysiological characteristics better predicts the performance of unseen users (error = 9.3% and correlation between online performances and predicted performances = 0.51 with $p = 0.01$) than using all user-related characteristics (error = 9.61% and correlation between online performances and predicted performances = 0.43 with $p = 0.04$). However, the model combining neurophysiological and traits characteristics revealed interesting results. Indeed, the mental rotation score had a negative weight in this model while different studies [52, 65] revealed a positive correlation between the mental rotation and online BCI performances. Therefore, this observation reinforces the hypothesis that the mental rotation score and therefore spatial abilities

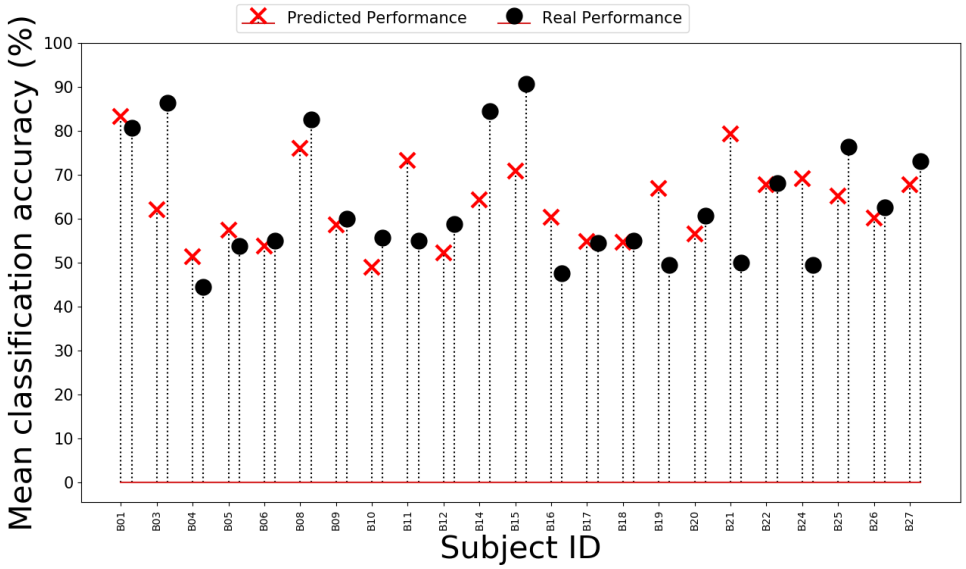


Figure 7.4: Result showing predicted performance in red (cross) of each subject of Dataset 2B generated using the model defined using Dataset 2A. In black (circle), the real performance of each subject.

might be related to BCI performances, however, the direction of the impact remains unclear. In addition, the tension factor, present in the between-subject model of Dataset 3A (see Chapter 5) was also present with a negative weight in this model. Subjects with higher tension level tend to overflow with energy and thus can become impatient [126]. Note that the tension factor is included in the 16pf5 global factor anxiety, which has already been shown to be important in predicting BCI performances (Chapter 5). Two other factors of personality i.e., reasoning (with a positive weight) and openness to change (with a negative weight) were selected in our model. High reasoning scorers have good reasoning skills and high openness to change scorers have a tendency to think about how to improve things and take advantage of experience [126]. All together, those results reveal that anxiety, the ability to solve problems but also the desire to create and experience novelty might be related to BCI performances.

Concerning neurophysiological characteristics extracted from EO and EC baselines, the presence of α-band power related factors (α-band power (EO and EC)) could suggest that the higher the difference between the α-band power at rest and the one during an MI task , the more likely it is to successfully control a MI-BCI, which is in line with previous studies [14,177,178]. Note that the SMR predictor (EO) was not selected in this model which could suggests that the α-band power at rest might have a stronger connexion with BCI performances than the SMR predictor. Another reason could be that a user trait factors are related to the SMR predictor and therefore not included in the model. Indeed, our results revealed that the SMR predictor was correlated to the apprenhension factor, itself correlated to the dominance factor which was correlated to the tension factor, included in the model. Finally, the exponent parameter λ of the $1/f^{\lambda}$ noise spectrum (i.e., its slope) was again selected with a strong positive weight in our model as well as the θ-band power in EC resting state condition, selected with a strong negative weight. This result reinforce the conclusions made in Chapter 6 i.e., that the functional relevance of the 1/f power and the θ-band power in resting state neural activities substantiate the necessity of isolating the 1/f component from oscillatory activities when studying the functional relevance of spontaneous brain activities.

Finally, as a sanity check, we tried to create a computational model of MI-BCI performances using Dataset 2A and 2B and only users' traits as predictive factors but our results were not conclusive which strengthens the idea that users' trait only cannot predict MI-BCI performances.

Conclusion

For this second contribution of our work, we suggested new predictors of MI-BCI and proposed computational models of BCI user training that could predict the performances of unseen BCI user based on their neurophysiological characteristics or their neurophysiological characteristics combined with stable characteristics such as traits and skills. Those studies represent a first step in the understanding of MT-BCI. Indeed, we were able to show that personality factors related to anxiety but also cognitive skills such as spatial abilities could be related to user performances. In addition, neurophysiological characteristics extracted from open and closed eyes resting state baselines provide information about MI-BCI performances of unseen users. Those characteristics are related to the α- and θ-band powers at rest but also signal to noise ratio.

To go deeper in the understanding of MT-BCI learning, in addition to the characteristics directly related to the users, we consider that it would be interesting to also consider characteristics extracted from the machine-learning methods used to build the classifier. Indeed, in BCI systems, the user learns directly from the feedback, which itself is directly related to the classification methods used. Therefore, such analysis could provide additional information to improve user learning. The next Part is dedicated to the study of the machine learning results to predict MI-BCI performances and therefore better understand learning. In a fist chapter, i.e. Chapter 8 we will focus on one specific data-driven method used during the system calibration i.e., the most discriminant frequency band algorithm. In a second chapter, i.e., Chapter 9 we will introduce a preliminary study of two famous algorithms used in the MT-BCI community, the common spatial pattern (CSP) algorithm and the linear discriminant analysis (LDA) classifier.

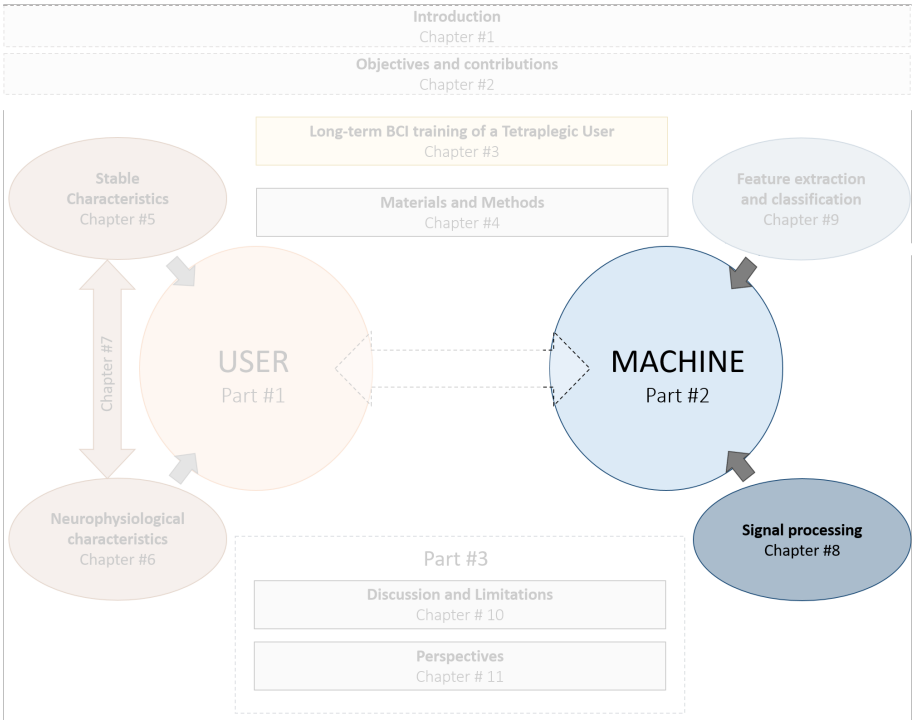
Part II

Machine-related Characteristics

8

Optimization algorithm: The Most discriminant frequency band selection

ROADMAP -



QUICK SUMMARY -

This chapter studies the relationship between BCI performances and characteristics of the subject-specific Most Discriminant Frequency Band (MDFB) selected by a popular heuristic algorithm.

- First, we studied the relationships between MI-BCI performances and subject-specific frequency band characteristics. Our results showed a correlation between the selected MDFB characteristics (mean and width) and performances.
- Then, we investigated a possible causality link. To do so, we compared, online, performances obtained with a constrained (enforcing characteristics associated to high performances) and an unconstrained algorithm. Although we could not conclude on causality, average performances using the constrained algorithm were the highest.
- Finally, to understand the relationship between MDFB characteristics and performances better, we used machine learning to 1) predict MI-BCI performances using MDFB characteristics and 2) select automatically the optimal algorithm (constrained or unconstrained) for each subject. Our results revealed that the constrained algorithm could improve performances for subjects with either clearly distinct or no distinct EEG patterns.

Contents

COLLABORATORS -

Maria Sayu Yamamoto

RELATED PAPER/CONFERENCE PAPER -

BENAROCH C., JEUNET C. & LOTTE F., (2021). MI-BCI Performances correlate with subject-specific frequency band characteristics. *In BCI 2021-8th International Meeting of the Brain-Computer Interface Society.*

BENAROCH C., JEUNET C. & LOTTE F., (2021). When should MI-BCI feature optimization include prior knowledge, and which one? *Brain-computer interfaces: Papers from the Virtual Brain-Computer Interface Meeting - accepted with major revisions.*

8.1 Introduction

During the calibration phase of an MI-BCI system, discriminative data-driven learning methods are commonly used to perform EEG feature extraction. One popular signal processing technique for EEG-based MI-BCIs is the Common Spatial Pattern (CSP) algorithm, which learn spatial filters that best discriminate between two MI classes [17, 18]. CSP filters maximize the variance of the spatially filtered signal under one condition while minimizing it for the other class. Before using CSP filters, several parameters have to be selected: the band pass filter, the time interval and the number of filters to use. When calibrating the system, it is common to use general settings (i.e, the same settings) for all subjects, e.g., the standard 8-30 Hz pass-band for MI-BCIs [17]. Yet, individually choosing the hyperparameters for CSP-based classification could improve online performances [19]. Some data-driven methods have been developed to select subject-specific hyperparameters [18]. Once the discriminative features identified, linear classifiers (e.g., LDA...) are most often used to distinguish classes [82].

Although these methods are commonly used and have proven to be effective, they are almost exclusively data-driven. They include very little neurophysiological prior, and rather trust the (potentially noisy) EEG data recorded during the calibration phase. We hypothesise that, to be more effective, these methods should take into consideration some constraints that may need to be identified. Indeed, are all properties of the features or classifiers learned from BCI data equally likely to be associated with good performances in practice? If not, what properties are more often associated with superior decoding performances? Would enforcing such properties in machine learning BCI algorithms lead to better decoding performances in practice?

In this chapter, we aim at answering this research question for a specific machine learning BCI algorithm: the frequency band selection algorithm of Blankertz et. al [18]. This heuristic algorithm selects the frequency band the power of which, in the sensorimotor channels, maximally correlates with the class labels. While often effective in practice, the lack of constraints in this algorithm may lead to the selection of frequency bands that are suboptimal. For instance, for illustrative purpose, this algorithm could select a most discriminant frequency band that is only 0.5 Hz wide whereas one may wonder whether such a narrow - and thus probably too specific - band is likely to lead to high decoding performances. Thus, in this chapter, we study the impact of the properties of the most discriminant frequency band (MDFB) selected by machine learning algorithms on online performances, with three different analyses. First, we investigated possible relationships between the characteristics of this selected MDFB and online BCI performances. Then, in order to study a possible causality link between these characteristics and performances, we designed an experiment using a new version of the algorithm with added constraints to enforce characteristics associated with superior performances. We compared online BCI performances obtained using the constrained algorithm with those of subjects using the unconstrained (i.e., original) algorithm. Finally, to better understand the relationships between the characteristics of the MDFB and online BCI performances, and to find out whether and if so when to use constraints, we used a LASSO (least absolute shrinkage and selection operator) [125] regression to determine explanatory and predictive models of MI-BCI performances based on MDFB characteristics. We then used a classification algorithm (here a decision tree) to automatically select the best algorithm (i.e. constrained or unconstrained) for each subject, based on the factors identified by these models.

8.2 Studying the relationships between MI-BCI performances and subject-specific frequency band characteristics

8.2.1 Objective

The frequency band selection algorithm is a data-driven method that selects the frequency band the power of which, in the sensorimotor channels, maximally correlates with the class labels, with little consideration for the resulting human performances and with little constraints based on prior knowledge. In this section, in order to determine if such an approach could benefit from using prior knowledge, we first investigated the relationships between the subject-specific frequency bands selected by the algorithm and users' online BCI performances. Our first objective was to determine if there was a correlation between characteristics of the selected band and online performances. Then, based on those first results, we aimed at identifying and proposing constraints to add to the original frequency band selection algorithm, in order to exploit this new knowledge as prior, and thus possibly improve the approach.

8.2.2 Materials and Methods

8.2.2.1 Dataset

We used the data of fifty-five subjects (N=55) out of the sixty subjects from the experiment in [130] (29 women; age 19-59; $\bar{X}$ =29; SD=9.318). This dataset is Dataset 2A. We discarded five subjects due to technical issues during the BCI recording session (e.g., a missing calibration run, or a mislabelling of EEG channels that resulted in an erroneous calibration).

8.2.2.2 The frequency band selection algorithm

In order to discriminate the MI tasks from EEG data, participant-specific spectral and spatial filters have been optimized on the data of the first two runs (calibration runs). Before computing spatial filters to build the classifier, the participants' MDFB was selected using the heuristic algorithm proposed in [18] (see Algorithm 1). It selects the frequency band the power of which, in the sensorimotor channels, maximally correlates with the class labels. More precisely, the EEG trials were first spatially filtered using a Laplacian filter around C3 [$4C3 - FC3 - CP3 - C5 - C1$] and C4 [$4C4 - FC4 - CP4 - C6 - C2$]. Then, for each Laplacian channel and each frequency, the correlation coefficient between the log-transformed band power of the Laplacian channel signal and the class labels was computed across trials. The frequency for which the sum of the two Laplacian channel correlation coefficients was maximal was used as a reference. In the case where the correlation coefficient for this frequency of reference was negative in a Laplacian channel, all the signs of the correlation coefficient were reversed. The frequency for which the sum of correlation coefficients across the two Laplacian channels was maximal was selected and named $fmax$. Finally, the frequency band was enlarged around $fmax$ until the correlation signal (named $fscore$) exceeded a threshold of $fscore(fmax) \times 0.05$.

8.2.2.3 Variables and factors

The metric used for online performance was the the online trial-wise accuracy, i.e. the default accuracy measure provided online in the MI-BCI scenarios of OpenViBE, which is also the performance metric that the experimenters were seeing online. We explored whether this classification accuracy value (dependant variable) correlated with two characteristics of the MDFB (independent variables), namely its mean frequency and width, the computation of which is explained below.

The optimization algorithm presented previously provided us with the MDFB boundaries f_{min} and f_{max}. From these two values, for each subject, we computed the mean value (MDFB mean) and the width (MDFB width) of the selected MDFB:

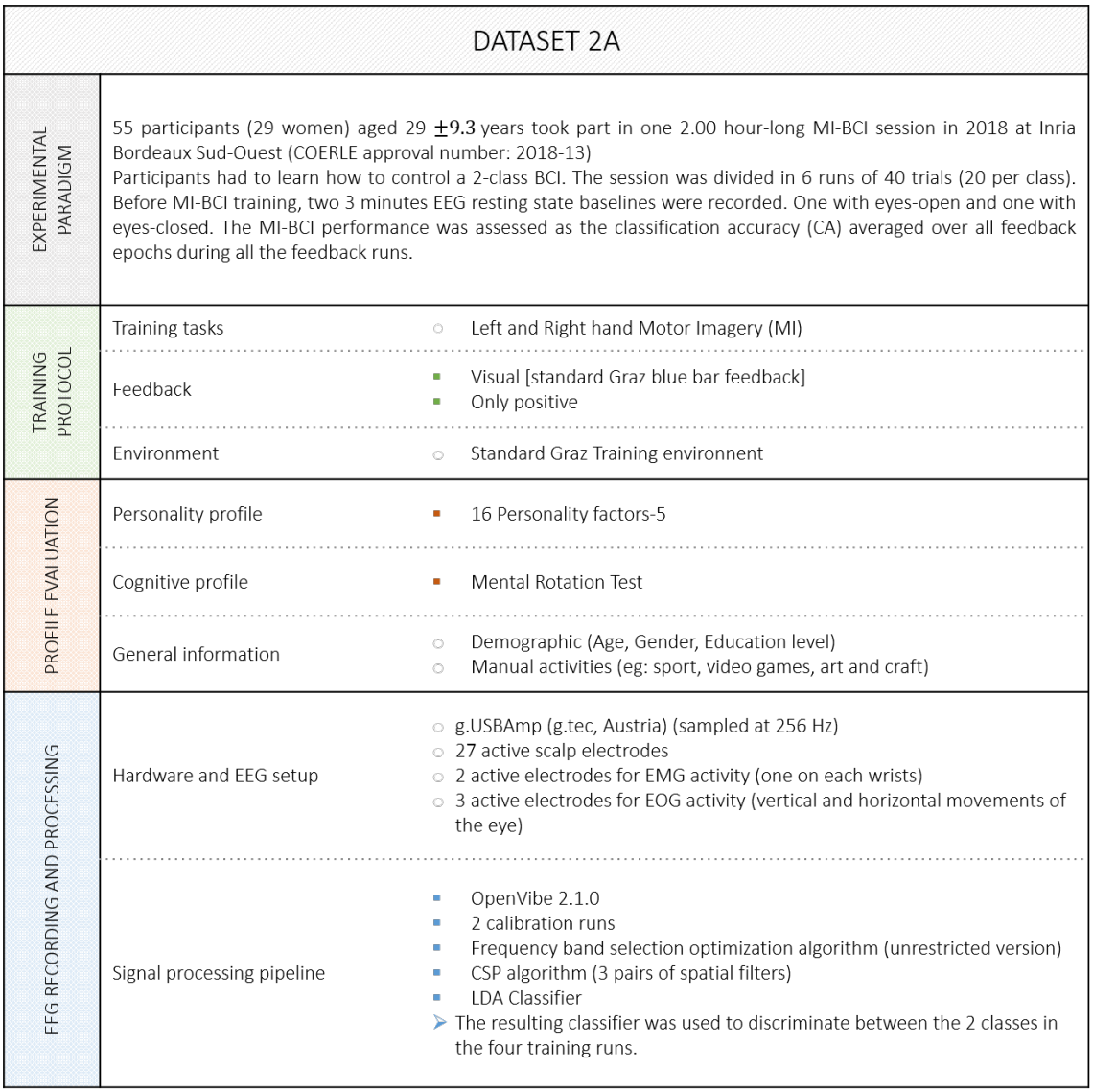


Figure 8.1: Summary table for dataset 2A.

– MDFB mean = $\frac{f_{min}+f_{max}}{2}$

– MDFB width = f_{max} - f_{min}

8.2.2.4 Analyses

None of the three variables (namely the MDFB mean, width and the MI-BCI performance) followed a normal distribution. Thus we used non-parametric tests:

– A Spearman's rank correlation test to measure the relationship between the MDFB mean or width and MI-BCI performances.

– A Wilcoxon rank sum test to compare the MI-BCI performances of two groups of subjects with different MDFB mean or width (i.e., one group below a threshold and the other above) in order to identify possible constraints to add to a new improved version of the algorithm. The thresholds were fixed visually using the bivariate distributions of the mean online classification accuracy and the MDBF width and mean (see Figure 8.2) so that both groups were about the same size.

Algorithm 1 Unconstrained Algorithm, i.e., original frequency band selection algorithms from Blankertz et. al [18]

Let $X_{(c,i)}$ denote trial i at channel c with label y_i and let C denote the set of channels.
$dB_c(f,i) \leftarrow$ log band-power of$X_{(c,i)}$at frequency f (f from 5 to 20 Hz)
$score_c(f) \leftarrow \text{corrcoef}(dB_c(f,i), y_i)_i$
$f_{max} \leftarrow argmax_f \sum_{c \in C} score_c(f)$
$score^*_c(f) \leftarrow \begin{cases} score_c(f) & \text{if } score_c(f_{max}) > 0 \\ -score_c(f) & \text{otherwise} \end{cases}$
fscore(f) $\leftarrow \sum_{c \in C} score^*_c(f)$
$f^*_{max} \leftarrow argmax_f fscore(f)$
$f_0 \leftarrow f^*_{max} f_1 \leftarrow f^*_{max}$
while $fscore(f_0 - 1) \geq fscore(f^*_{max}) * 0.05$ **do**
    $f_0 \leftarrow f_0 - 1$
while $fscore(f_1 + 1) \geq fscore(f^*_{max}) * 0.05$ **do**
    $f_1 \leftarrow f_1 + 1$
Return frequency band$[f_0; f_1]$

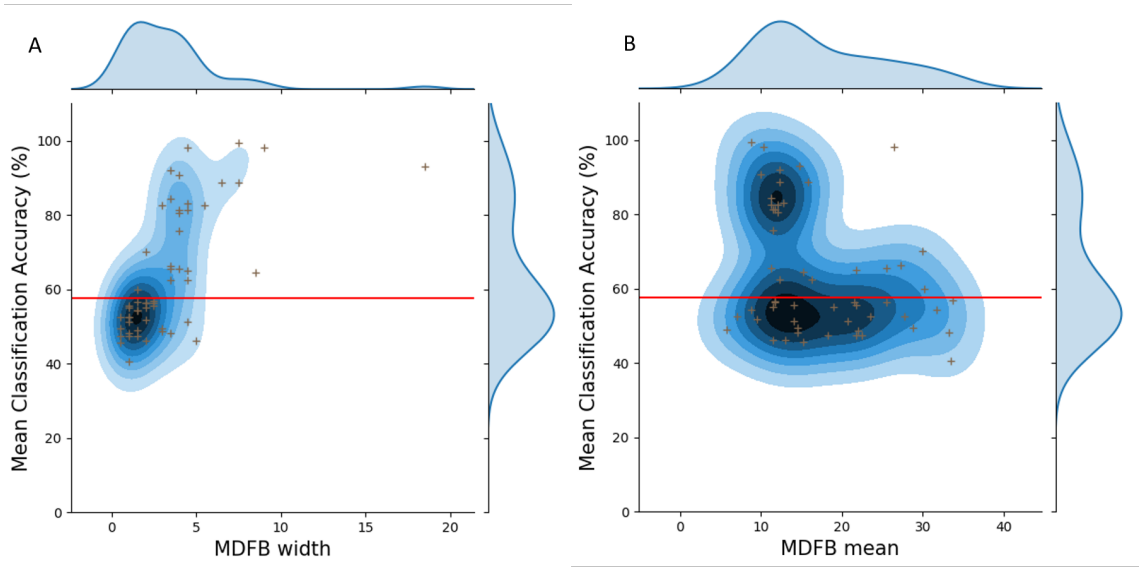


Figure 8.2: Bivariate distribution of the the mean online classification accuracy and (a) the frequency band width and (b) the mean frequency of the band. In red, the statistical chance level for 2-classes, 80 trials per class and $\alpha = 5\%$ [122]

8.2.3 Results

First results showed correlations between performances and the two characteristics of the selected frequency band. A significant negative correlation was observed between MI-BCI performances and the *MDFB mean* frequency (r=-0.28, p $\leq$ 0.01) and a significant positive one between MI-BCI performances and the *MDFB width* (r=0.70, p $\leq$ 0.01).

Figure 8.2 represents the bivariate distribution of the *MDFB width* and the *MDFB mean* with the MI-BCI performances, i.e., the mean classification accuracy.

Other analyses revealed that:

- Subjects with an *MDFB width* lower than 3.5Hz have lower MI-BCI performances compared to subjects with a larger *MDFB width* (U=-4.9, p<10-5) (see Figure 8.3 A).
- Subjects with an *MDFB mean* value above 16Hz seem to have lower performances than subjects with an *MDFB mean* value under 16Hz (U=2.2, p<0.05) (see Figure 8.3 B).

Thresholds (i.e., 3.5Hz and 16Hz) were fixed visually using Figure 8.2 results.

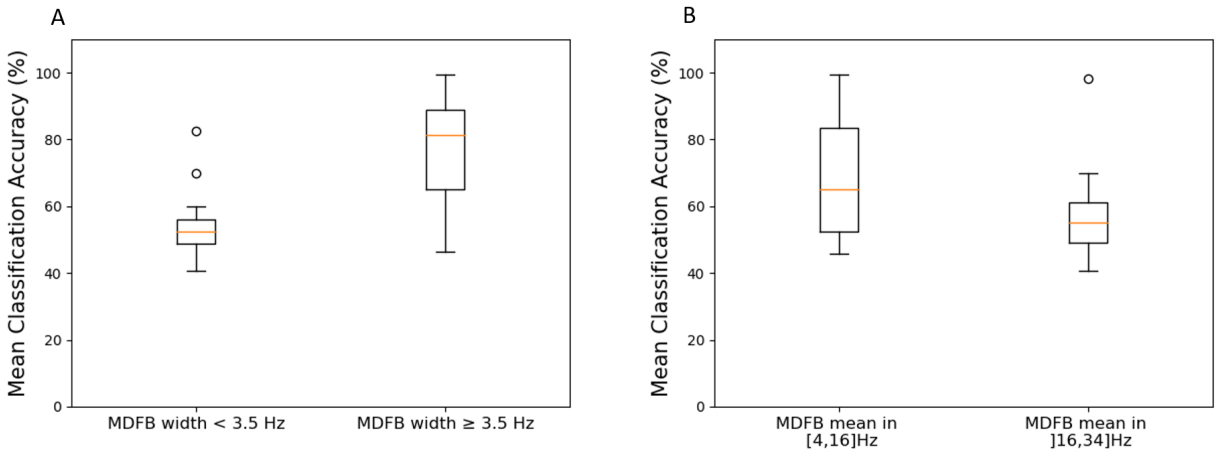


Figure 8.3: Boxplot that show (A) a significant difference of the median of the two groups (*MDFB mean* value in [4,16] or in]16,35] (U =2.2, p<0.05) and (B) a significant difference of the median of the two groups (*MDFB width* <3.5 or *MDFB width* ≥3.5 (U =-4.9, p<e-5)

8.2.4 Intermediate Discussion

These results suggest that there is a correlation between the selected user-specific frequency band characteristics (i.e., the mean value and the width of the MDFB selected) and the classification accuracy. The user-specific frequency bands were selected within the α-β range ($5-35$Hz) by using the heuristic proposed by Blankertz et al. in [138]. Results also suggest that (1) subjects with MDFB Length $<$ 3.5Hz seem to experience difficulties controlling a BCI compared to subject with higher *MDFB width*; (2) subjects with an *MDFB mean* value above 16Hz (in the β band) seem to have lower performances than subjects with an *MDFB mean* value under 16Hz (in the α band). However, this correlation does not imply any causal relationship between those factors. A selected user-specific frequency band in α-lowβ could cause higher performance, or users with higher performance could modulate their user-specific frequency bands in α-lowβ range, or a third factor, not assessed here, could have impacted both the user-specific frequency band and performances. Therefore, we cannot conclude on causal effects from this study alone.

The next step in our work would thus be to determine if there is a causal link between both variables and, if any, the direction of this link. In the following section, we thus investigated how manipulating the MDFB by restricting the selection of frequency bands affected BCI performance, to assess possible causal effects.

8.3 Investigating the causal relationship between MI-BCI performances and subject-specific frequency band characteristics

8.3.1 Objective

In order to investigate a possible causation between BCI performances and characteristics of the MDFB, we designed a constrained version of the MDFB selection algorithm that enforced i) an *MDFB width* larger than 3.5Hz and ii) an *MDFB mean* value under 16Hz. We used it in an online experiment specifically designed for this study (Dataset 2B). We then compared the online performances obtained for subjects in this experimental group, Dataset 2B (using the constrained algorithm) with online performances of matched subjects in a control group, named Dataset 2A*. These subjects were selected from Dataset 2A and were then provided with the unconstrained, i.e., original algorithm.

8.3.2 Materials and Methods

8.3.2.1 Constrained version of the frequency band selection algorithm

In order to enforce (1) an *MDFB width* larger than 3.5Hz and (2) an *MDFB mean* value under 16Hz in the MDFB selection algorithm, we did the following:

(1) We smoothed, using a Savitzky-Golay filter [198], for each set of Laplacian channels, the $scorec(f)$ signal (Algorithm 2, line 6), which represents the correlation, at each frequency, between the band power of each trial and its class label. Smoothing the signal avoids having high isolated peaks in the scorec(f) signal which had the consequence of leading to narrow *MDFB width*. Then, we proceeded as in the original algorithm. If the *MDFB width* was still below 3.5Hz, we increased the band on the side where the sum of the the two channel correlation coefficients was the highest until reaching the desired width, i.e., at least 3.5Hz (Algorithm 2, lines 13-20).

(2) We constrained the search of the frequency (Algorithm 2, lines 22-26) for which the sum of the two Laplacian channel correlation coefficients was maximal (i.e., f^*_{max}) to frequencies between 8 and 16Hz (i.e., in α-lowβ range) and we used the algorithm for frequencies restricted to be between 5 and 20Hz to avoid having an MDFB mean value above 16Hz.

Algorithm 2 Constrained Algorithm

1: Let $X_{(c,i)}$ denote trial i at channel c with label y_i and let C denote the set of channels.
2: $dB_c(f,i) \leftarrow$ log band-power of$X_{(c,i)}$at frequency f (f from 5 to 20 Hz)
3: $score_c(f) \leftarrow$ corrcoef$(dB_c(f,i), y_i)_i$
4: $f_{max} \leftarrow argmax_f \sum_{c\in C} score_c(f)$
5: $score^*_c(f) \leftarrow \begin{cases} score_c(f) & \text{if } score_c(f_{max}) > 0 \\ -score_c(f) & \text{otherwise} \end{cases}$
6: fscore(f) $\leftarrow$ savgolFilter($\sum_{c\in C} score^*_c(f)$)
7: $f^*_{max} \leftarrow argmax_f fscore(f)$
8: $f_0 \leftarrow f^*_{max} f_1 \leftarrow f^*_{max}$
9: **while** $fscore(f_0 - 1) \geq fscore(f^*_{max}) * 0.05$ **do**
10: $\quad f_0 \leftarrow f_0 - 1$
11: **while** $fscore(f_1 + 1) \geq fscore(f^*_{max}) * 0.05$ **do**
12: $\quad f_1 \leftarrow f_1 + 1$
13: **while** $fscore(f_1) - fscore(f_0) < 3.5$ **do**
14: $\quad$ **if** $fscore(f_1) < fscore(f_0)$ **then**
15: $\quad\quad f_0 \leftarrow f_0 - 1$
16: $\quad$ **else if** $fscore(f_1) > fscore(f_0)$ **then**
17: $\quad\quad f_1 \leftarrow f_1 + 1$
18: $\quad$ **else if** $fscore(f_1) = fscore(f_0)$ **then**
19: $\quad\quad f_0 \leftarrow f_0 - 1$
20: $\quad\quad f_1 \leftarrow f_1 + 1$
21:
22: **if** $\frac{f_0+f_1}{2} > 16$ or $\frac{f_0+f_1}{2} < 8$ **then**
23: $\quad$ **while** $\frac{f_0+f_1}{2} > 16$ **do**
24: $\quad\quad f_0 \leftarrow f_0 - 1$
25: $\quad$ **while** $\frac{f_0+f_1}{2} < 8$ **do**
26: $\quad\quad f_1 \leftarrow f_1 + 1$
27: Return frequency band$[f_0; f_1]$

8.3.2.2 Datasets

Dataset 2B: The experimental group:

We included twenty-five subjects in this experimental group with the same inclusion criteria and experimental protocol (instructions, tasks, feedback and training environment) as Dataset 2A in order to match this dataset as closely as possible to Dataset 2A (see Figure 8.4. The only difference was the algorithm that we used to select the MDFB (see Algorithm 8.3.2.1).

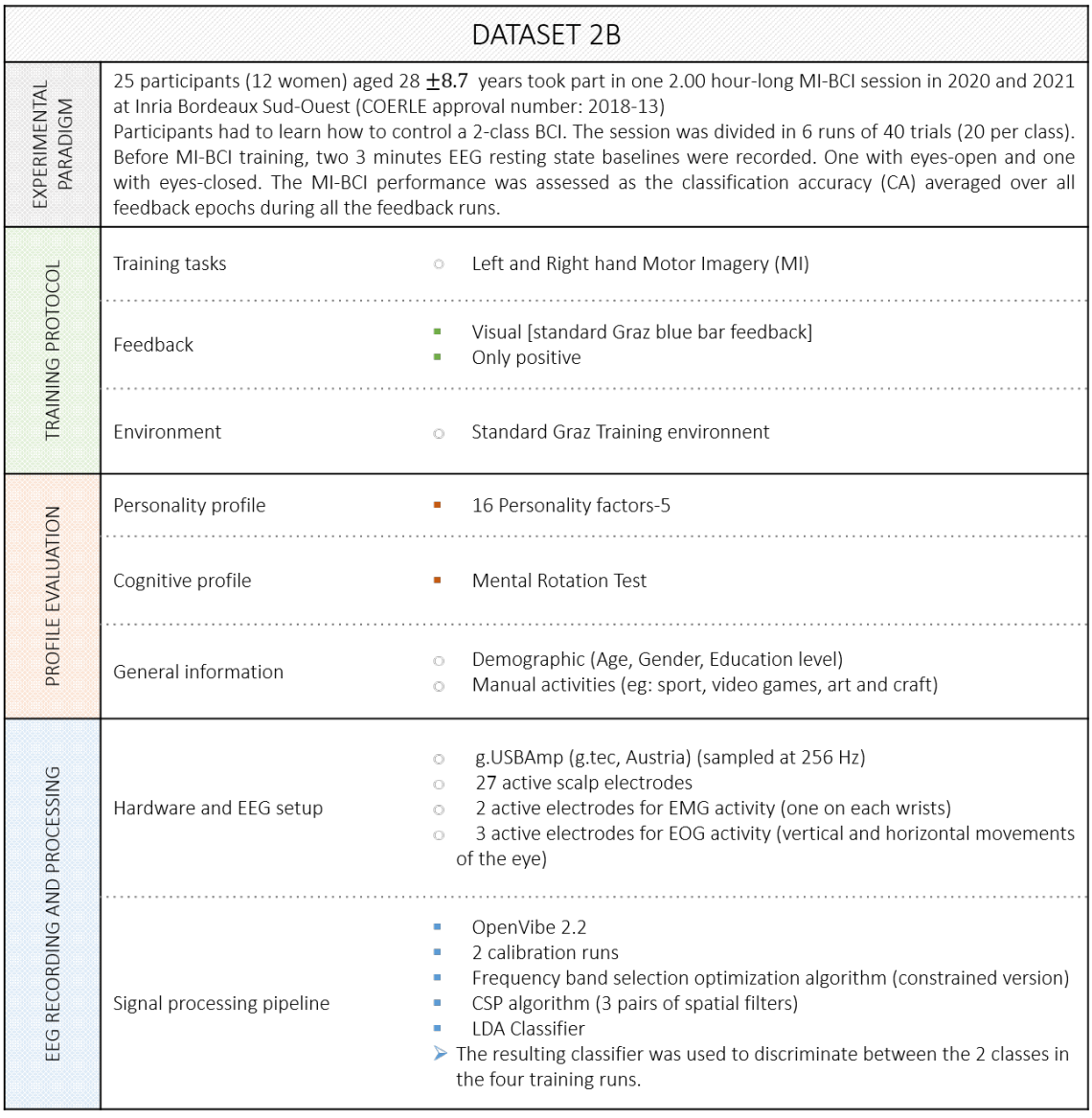


Figure 8.4: Summary table for dataset 2B.

Dataset 2A*: The control group:

As the objective was to study the effect of the new constrained algorithm on online performances, we paired subjects from Datasets 2A and B as follows. We first obtained, offline, the MDFB mean and width of subjects from Dataset 2B using the unconstrained (i.e., original) algorithm. Then, each subject was paired with a subject from Dataset 2A with similar MDFB mean and width. This

Experimental Group (Dataset 2B)						Control Group (Dataset 2A*)			
ID	Constrained Algorithm			Unconstrained Algorithm		ID	Unconstrained Algorithm		
	MDFB Mean (Hz)	MDFB Width (Hz)	Classification Accuracy (%)	MDFB Mean (Hz)	MDFB Width (Hz)		MDFB Mean (Hz)	MDFB Width (Hz)	Classification Accuracy (%)
B1	10	8	80.6	11.75	4.5	A43	11.75	4.5	81.3
B2	12	5	86.3	25.5	15	A44	25.5	4	65,6
B3	12	4	44.4	12	2	A06	11.5	2	46.3
B4	10.25	10.5	53.8	18.25	0.5	A10	18.25	0.5	47.5
B5	11.75	3.5	55	11.5	1	A52	11.5	1	55
B6	12.25	8.5	97.5	12.5	8	A23	12.25	6.5	88.8
B7	12.25	3.5	82.5	12.25	1.5	A48	11.75	1.5	56.6
B8	12.25	3.5	60	11.25	1.5	A24	8.75	1.5	54.4
B9	13.25	6.5	55.6	20.25	4.5	A40	20.75	4.5	51.3
B10	13.25	13.5	55	26.75	3.5	A43	27.25	3.5	66.3
B11	8.75	3.5	58.8	33	3	A02	33.25	3.5	48.1
B12	10.75	9.5	84.4	13.25	4.5	A01	12.25	4.5	62.5
B13	10.5	6	90.6	11	4	A13	11.5	4	75.6
B14	15	8	47.5	29	1	A22	28.75	0.5	49.4
B15	9.75	3.5	54.4	21.75	1.5	A25	21.75	1.5	47.5
B16	12	15	55	32	6	A29	26.5	9	98
B17	13.5	11	49.4	17	2	A03	19	2	55
B18	14.75	3.5	60.6	20.5	3	A11	22	3	48.8
B19	13.75	7.5	50	29	2	A30	30	2	70
B20	9.5	8	68.1	11.5	5	A19	13	5	46.5
B21	12	10	71.9	12.25	5.5	A21	11.25	5.5	82.5
B22	9.75	3.5	49.4	8	3	A38	9.75	2	51.9
B23	9.5	5	76.3	9.25	4.5	A42	11.75	4.5	81.3
B24	10.75	5.5	62.5	21.38	4	A51	21.75	4.5	65
B25	8.75	3.5	73.2	8.75	2.5	A15	11.75	2.5	56.25

Table 8.1: Data for experimental and control groups

enabled us to compare online the effects, if any, of the constrained algorithm (from Dataset 2B) for pairs of subjects with similar MDFB with the original algorithm. In the end, in Dataset 2A* we included 25 subjects (12 women; age: 20-57 year-old; $\bar{X}$ =29; SD=10). Table 8.1 shows selected MDFB characteristics and the mean classification accuracy for subjects in Dataset 2B and Dataset 2A*.

8.3.2.3 Analyses

Our objective was to compare the MI-BCI performances obtained with a constrained vs. unconstrained MDFB algorithm. Therefore, we selected in Dataset 2A*-B only the participants whose MDFB characteristics differed in both conditions. The following heuristics were used to select the participants:

$$\left| \frac{m_{constrained}}{m_{unconstrained}} - 1 \right| \leq 0.10 \tag{8.1}$$

where $m_{constrained}$ is the MDFB mean selected by the constrained algorithm and $m_{unconstrained}$ is the MDFB mean selected by the unconstrained algorithm.

$$\frac{d}{(l_{constrained} + l_{unconstrained}) - d} \leq 0.60 \tag{8.2}$$

where $l_{restricted}$ is the *MDFB width* selected by the constrained algorithm, $l_{unrestricted}$ is the *MDFB width* selected by the unconstrained algorithm, and d is the width of the range of frequencies that are common to the MDFB obtained by both algorithm (e.g., if the unconstrained algorithm selected 11-19 Hz while the constrained algorithm selected 8-13Hz, their common frequency range was 11-13Hz and thus $d = 2$ Hz). As a result, the first analyses are based on the results of 20 participants out of 25 for each group (B06, B13, B21, B23 and B25 were removed as their MDFB characteristics were similar with both algorithms as well as A23, A13, A21, A42 and A15 as they were associated to B06, B13, B21, B23 and B25). As the classification accuracy of all the different groups did not follow a normal distribution, we used a Wilcoxon signed-rank test to compare the MI-BCI classification accuracy between the two algorithms.

8.3.3 Results

While the average online performances for the constrained algorithm are higher (M=60.7 ± 12.9) than for the unconstrained algorithm (M=58.3 ± 13.3), our analyses showed that this difference is not significant (W=67.5, p=0.13). Figure 8.5 illustrates our results.

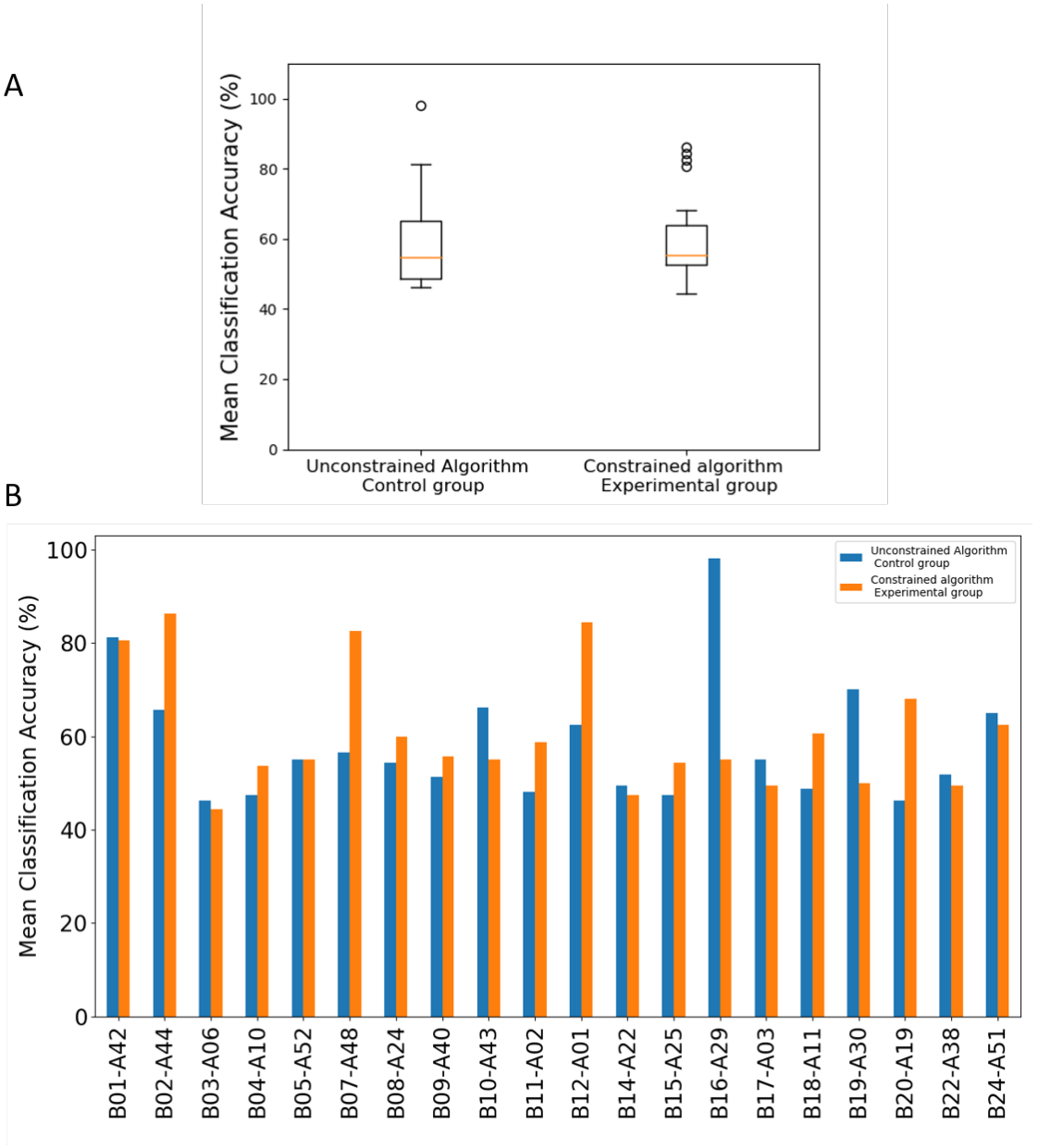


Figure 8.5: (A) Boxplot representing the distribution of the online classification accuracy (%) for the two groups (i.e., the experimental group using the constrained algorithm and the control group using the unconstrained algorithm, before using the CSP algorithm.) (B) Graph representing the online classification accuracy (%) for each subject.

8.3.4 Intermediate Discussion

These results suggest that there is no statistical difference in the classification accuracy when using the constrained or the unconstrained algorithm. The constrained algorithm was associated with similar or better performances than the unconstrained algorithm in most subjects (i.e., between -5% and 46% of improvement in 80% of the subjects), but also detrimental for 20% of the subjects (a drop between 10% and 43% in online performances). It is interesting to note that for one pair of subjects (B16-A29), adding constraints was strongly detrimental and that in this case, it seems that subject A29 was able to produce discriminant signals in the high β band (22–31 Hz). Three other pairs of subject had better performances with the unconstrained algorithm, which could be due to the fact that the *MDFB width* selected with the constrained algorithm was not specific enough, i.e., too broad (larger than 10 Hz). Indeed, in the constrained algorithm, we decided to smooth, for each Laplacian channel, the signal representing the correlation, at each frequency, between the band power of each trial and its class label. Although this prevented us from having narrow MDFBs, it resulted in increasing the *MDFB width* for several subjects by including most of the pre-determined maximal band (i.e. 5-20Hz) in the MDFB.

Finally, despite these four cases, modifying the algorithm by putting constraints led to better or similar online performances in most cases (N=14) . This observation may suggest that adding constraints could be a solution to help most subjects learn how to control a BCI but not all of them. Is it possible to find other characteristics based on the MDFB to predict online performances? Could we use these characteristics to automatically select the optimization algorithm to use during the calibration for each subject? These questions are investigated in the next section.

8.4 Gaining a better understanding of the relationship between MI-BCI performances and subject-specific frequency band characteristics

8.4.1 Objective

In the two previous sections, we showed that even though there was a correlation between the selected user-specific frequency band characteristics (i.e., the mean value and the width of the MDFB selected) and the classification accuracy, adding constraints to the frequency band selection algorithm did not statistically improve user performances, even though this led to similar or better performance for 78% of the subjects. The reasons for this variability being unclear, we led additional analyses in order to:

- identify new predictors of MI-BCI performances originating from both constrained and unconstrained algorithms.
- create a statistical model that can predict MI-BCI performances based on characteristics of the MDFB.
- create a statistical model to select the most relevant algorithm for each subject.

8.4.2 Materials and Methods

8.4.2.1 Datasets

For this study we will use Datasets A and B where the frequency band selection algorithm was used for online experiments, during the calibration before using the CSP algorithm. In Dataset 2A the unconstrained algorithm was used whereas in Dataset 2B, we used the constrained algorithm.

8.4.2.2 Variable and Factors

To identify new predictors, we used the signal representing the sum of the two correlation coefficients (of the two Laplacian channels) at each frequency, between the band power of each trial and its class label (i.e., $fscore$) used to select the MDFB.

Then, as the MDFB can only be chosen in one specific band, we modeled each $fscore$ curve as a Gaussian function g of the frequency f, by using non-linear least squares to fit the function g to $fscore$:

$$g(f) = ae^{\frac{-(f-f0)^2}{b}} \tag{8.3}$$

where the coefficient a represents the maximum value of the g curve, $f0$ the mean value of the Gaussian g and b is a representation of its variance.

In the case in which the optimization algorithm did not converge (18% of the subjects), we enforced default values for a (we used the maximum value of $fscore$ in the MDFB) and $f0$ (we used the MDFB mean) and used the optimization algorithm to obtain b.

Coefficient a allowed us to assess the $fscore$ variability within the MDFB. Indeed, the more stable and high the $fscore$ signal within the MDFB, the higher a. Coefficient $f0$ gave information about the most discriminant frequency and b about the $fscore$ signal distribution.

Coefficients a, $f0$ and b were used as predictors, together with the MDFB width and mean obtained using the algorithm used online (cf Figure 8.6).

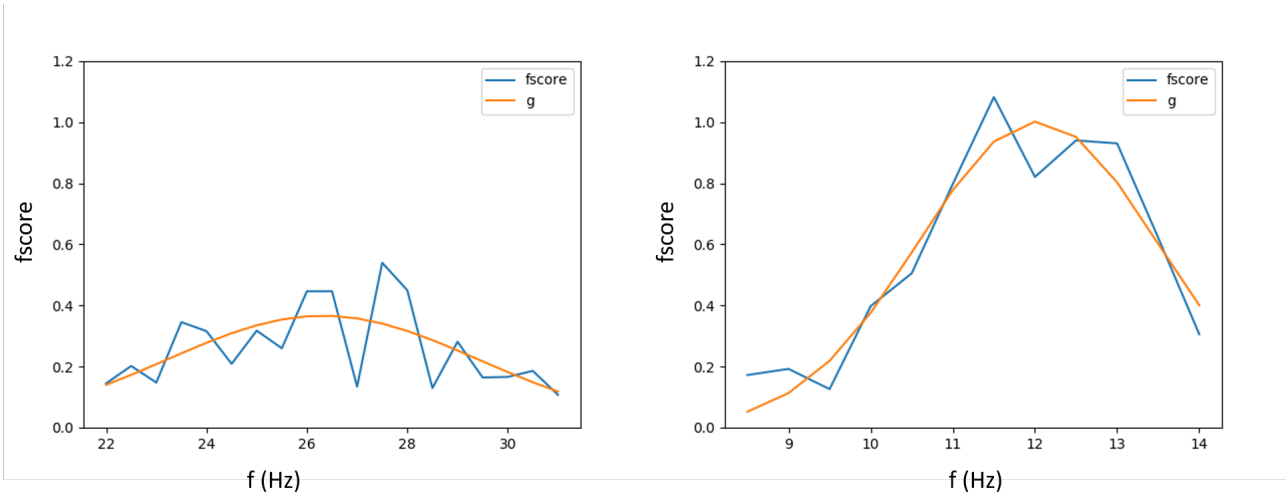


Figure 8.6: Two illustrations of the computation of the new performance predictors. The plots depict the $fscore$ of two participants (solid blue line) and the fitted values $g(f; a, f0, b)$ (solid orange line).

8.4.2.3 Analyses

Statistical model to predict/explain MI-BCI performances:

In order to build a Statistical model to predict/explain MI-BCI performances we used a LASSO [125] regression to predict the performances of MI-BCI users from the characteristics of the MDFB (see Chapter 4 for more details about LASSO regression). Table 8.2 represents a summary table of the method used to create our models of prediction.

STATISTICAL MODEL OF PREDICTION	
METHOD	LASSO
SUBJECTS	N=80 (76 after removing outliers, see Section Results)
EXPLANATORY FEATURES	▪ MDBF mean ▪ MDBF width ▪ A ▪ F0 ▪ b ➔ 5 features in total
VARIABLE TO PREDICT	The online trial-wise accuracy, i.e. the default accuracy measure provided online in the MI-BCI scenarios of OpenViBE
EVALUATION OF STABILITY AND RELIABILITY	▪ Leave-one-out cross validations ▪ Random model with 1000 permutations

Table 8.2: Summary table explaining the statistical model used to predict/explain MI-BCI performances.

Statistical model to select the most relevant algorithm for each subject

For this analyse, we used M=24 pairs of subjects from Datasets A* and B (see Table 8.1). Subject A52 paired with subject B05 were removed as their online performance were exactly the same in the two conditions (i.e., with constrained and unconstrained algorithm). This subject pair was thus not informative to predict which algorithm to use. In order to build a classification model to automatically select the most suitable algorithm for each subject, we used decision trees [199] to obtain models that could predict the algorithm (constrained or unconstrained) with which a user reaches the highest MI-BCI performances. We used a and $f0$ as predictive variable among the ones available before the use of the MDFB selection algorithm (i.e., a, $f0$, b) as according to our results below they were selected in 100% of the generated models. In order to evaluate the reliability of the different models, we used a leave-one-subject-out cross validation process. We also used an inner cross-validation (total number of training subjects: M-2) to automatically select hyper-parameters (i.e., the maximum depth of the tree and the minimum number of samples required to be at a leaf node). For each cross-validation, we compared the online performances of subjects from Dataset 2A* (all using the unconstrained algorithm) with performances obtained using the algorithm (either constrained or unconstrained) automatically selected by the decision tree model. We used the Wilcoxon signed-rank test to compare the two groups of performances. We also built a classification tree trained on all 20 subjects from Datasets A*, for interpretation purposes, in order to determine the critical values of a that could give information about which algorithm to use.

8.4.3 Results

We excluded from the analyses the subjects whose mean classification accuracy was above or below two standard deviations (SD) from the group's mean performance. Therefore, four subjects were removed from our analyses and we used data from 76/80 subjects.

8.4.3.1 Statistical model to predict/explain MI-BCI performances:

Our results indicated that the predictions made for Dataset 2A were better than chance with a mean absolute error $\sum_{i=1}^{N} \frac{|V_{pred(i)} - V_{real(i)}|}{N}$, n being the total number of models generated, of 10.48%. The mean absolute error (in %) of the random models were 11.60% and 12.20% for $p = .01$ and $p = .05$ respectively (see Chapter 4 for more details about the used method).

The *MDFB width* and the parameter a of the function g were selected in 100% of the generated models with a strong positive weight in comparison to the other factors' weights (see Figure 8.7). As a reminder, a represents the maximimun $fscore$ value of the fitted function g. Subjects with a higher parameter a and a higher *MDFB width* seem to obtain better performances than those with lower ones. In addition, the parameter $f0$ was selected in 100% of the generated models as well with a negative weight enhancing the hypothesis that subject with an MDFB mean value in the α-lowβ range tend to perform best. Indeed, as $f0$ represents the fitted Gaussian g mean value, its value should be close to the MDFB mean value.

We also computed the correlation between the real and predicted MT-BCI performances for each subject. We obtained a significant correlation i.e., $r = 0.46$ ($p = 3.10^{-5}$).

8.4.3.2 statistical model to select the most relevant algorithm for each subject:

The average online performances using the algorithm selected by the classification tree were similar (M=62.9 $\pm$ 15.2) to the average online performances using the unconstrained algorithm (M=62.3 $\pm$ 15.2) and our analyses showed that this difference did not reach significance.

Our final classification model (see Figure 8.8) suggests that subjects with an intermediate a value (between 0.4 and 0.8) would have higher online performances using the constrained algorithm whereas subjects with high a value have higher online performances using the unconstrained algorithm. Furthermore, subjects with a most discriminant frequency *f0* lower than 12.3Hz or higher than 18.9Hz have higher performances using the unconstrained algorithm whereas those with *f0* between 12.3 Hz and 18.9Hz would have higher performances with the constrained algorithm.

8.4.4 Intermediate discussion

Our analyses suggested that it was possible to predict the MI-BCI classification accuracy of users using parameters extracted from the optimization algorithms: the frequency band selection algorithms. The prediction models included two main factors: The *MDFB width* and the parameter a of the function g and one other factor with a negative smaller weight $f0$. Results from Section 8.2 already suggested that subjects with a higher *MDFB width* tend to perform better than subjects with a lower one. It also suggested that subjects with an MDFB mean value above 16 Hz (in the β band) seem to have lower performances, which is confirmed by the negative weight of $f0$ in all our models. Therefore, both those results indicate that the *MDFB width* and *mean* (or $f0$) could be a good predictor of MI-BCI performances. In addition, all models included the parameter a of the function g with a strong positive weight. As a represents the ability to produce strong and distinct EEG oscillations in the sensorimotor cortices, we believed that it could be interesting to explore the value of this parameter during the calibration of the system in order to refine the algorithm selection. We used parameter a and $f0$ to study the possibility of having a classification model that could select the most optimal algorithm (constrained or unconstrained) for each subject. Results showed that there is no statistical difference between predicted performances using the classification model (which select

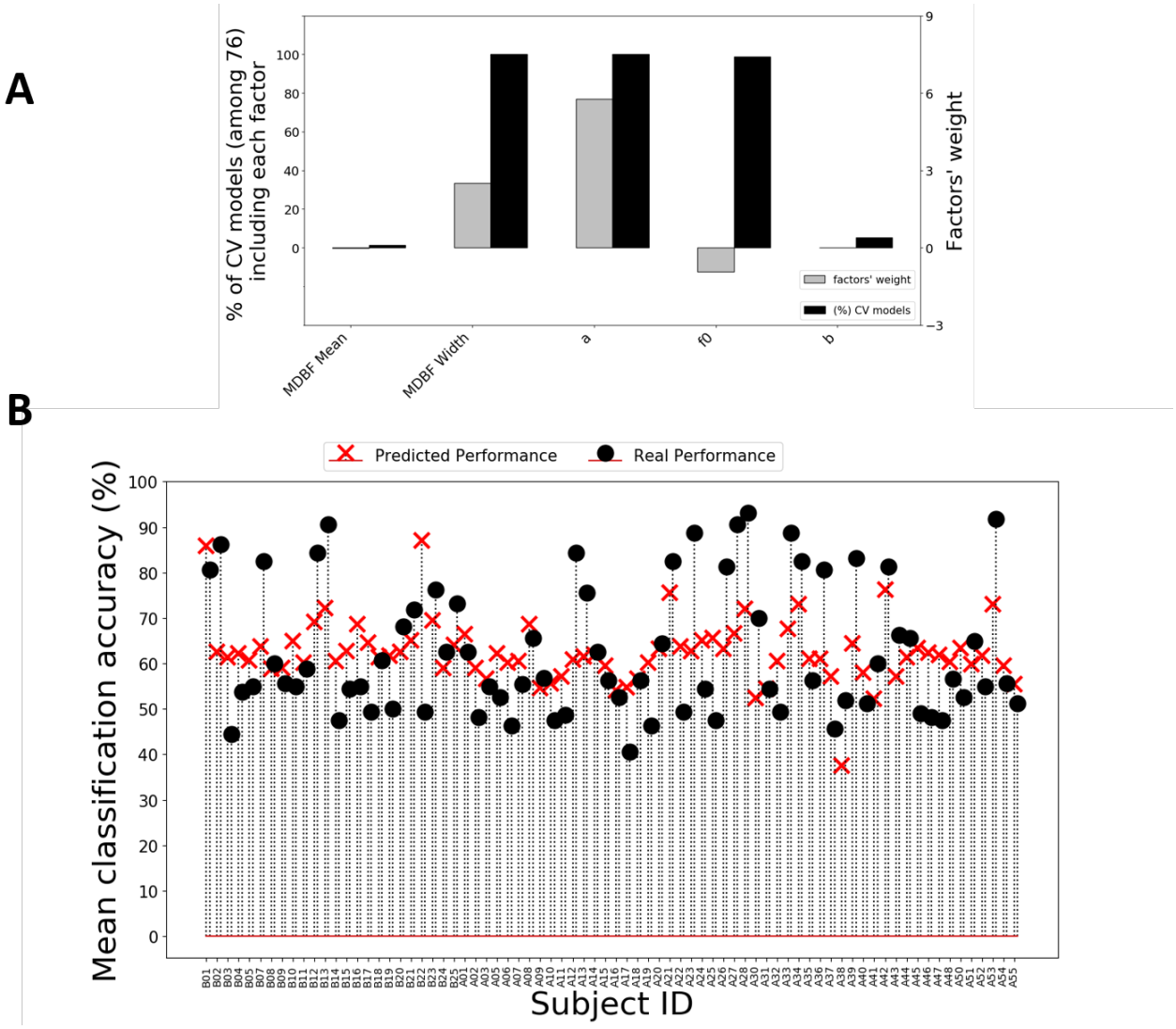


Figure 8.7: Results of the different models generated. (a) represents the percentage of Cross-Validation models including each factor (black) and their average weight (grey). (b) illustrates in black (circle), the real performance of each subject and in red (cross), the predicted performance of each subject generated using the model generated from the training dataset (All subjects except the target one).

the algorithm to use) and original performances (using the unconstrained algorithm). However, the mean performance while using the classification model (M=62.9 ± 15.2) was higher than the mean performance using the unconstrained algorithm (M=62.3 ± 15.2). The classification model suggest that for subjects with a high parameter a (i.e., subjects that produce strong and distinct signals) using the unconstrained algorithm could be a little detrimental (an average deterioration of 5.5%). When looking at the MDFB characteristics of those subjects we can observe that adding constraints increase the MDBF width by an average of 3Hz.

However, subjects with an intermediate a parameter could benefit from the added constraints (an average an improvement of 13%). For those subjects, adding constraints either selected an MDFB in the α low-β band while it was not the case with no constraints, or it widened the MDFB which might lead to a better learning of BCI control (at it may allow for more flexibility about which frequency to modulate, as compared to a narrow band). Finally, our results suggest that for subjects with a low parameter a (i.e., subjects that have difficulties producing strong and distinct signals) the most discriminant frequency $f0$ gives additional information to select the optimal algorithm. Indeed, for subjects with a $f0$ in the α-band (i.e., $< 12Hz$) adding constraints might not be useful

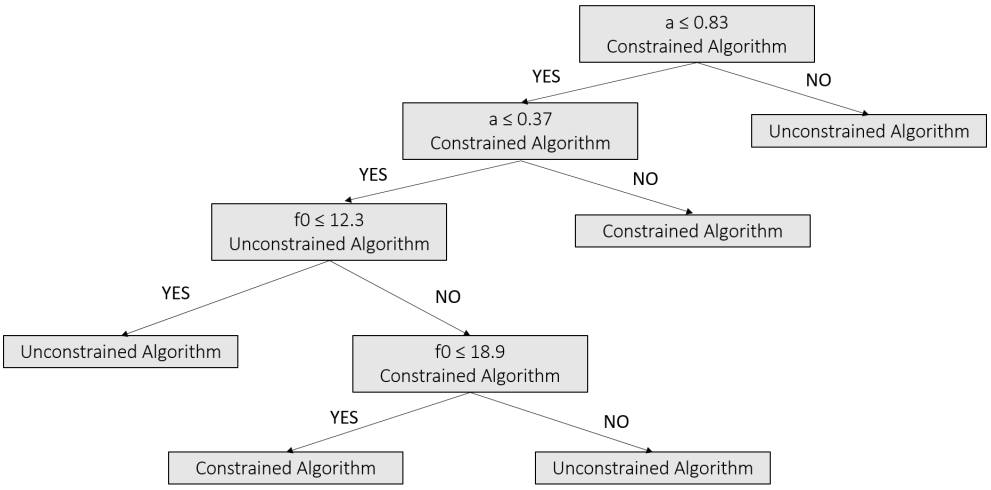


Figure 8.8: Representation of the decision tree.

(an average a deterioration of 2%). Subjects in this case have difficulties controlling the system with both algorithms. However, subjects with $f0$ in the low β band (between 12.3Hz and 18.9Hz) could benefit a little from the added constraints (an average an improvement of 6.5%). This may suggest that for those subjects with very weak SMR modulations and a $f0$ in the low β band, using default values (here our constraints), might be better than letting the classifier learn from data, as there is nothing clear in the data to learn from. However, for subject with very weak SMR modulations and a $f0$ in the high β band adding constraints is detrimental (an average deterioration of 10%), probably because the addition constraints prevents from having an optimal MDBF mean. Indeed, those subjects are able to produce discriminant signals in the high β band.

Finally, we find it important to note that our classification model was build on a small number of subjects (i.e., 24) and that our results do not allow us to conclude about a statistical improvement of online performances using the algorithm chosen by the model. Although our results do not allow us to conclude, we believe that such an investigation into data-driven BCI methods could be applied to other BCI algorithms, e.g., spatial filters, in the future, to better understand and possibly improve them.

8.5 Conclusion

In this study, we used the data from two different experiments to study the impact of the MDFB selected by an optimization algorithm during the calibration of an MI-BCI system. The objective was to determine if there was a relationship between characteristics of the chosen MDFB and subjects' online performances (i.e., classification accuracy), the causality of their relationship, if any, and if we could find a parameter that could be used to refine the algorithm to optimize it further. Our analyses suggested that there is a correlation between the selected user-specific frequency band characteristics (i.e., the mean value and the width of the MDFB selected) and the classification accuracy. To study a possible causality link between them, we added constraints to the algorithm, enforcing characteristics associated to higher performances (based on our first results), and used it in an experiment specifically designed for this study (**Dataset 2B**). We then compared the online performances obtained for subjects in this experimental group, **Dataset 2B** (using the constrained algorithm) with online performances of matched subjects in a control group Dataset 2A* from **Dataset 2A** (using the unconstrained, i.e., original algorithm). This study did not allow us to conclude on a causality link. Indeed, while average online performances obtained with the constrained algorithm were slightly higher than those obtained with the unconstrained algorithm, this difference did not reach significance.

Finally, in a last part we were able to predict the MI-BCI classification accuracy of users using parameters extracted from the MDFB optimization algorithms. To build our models, we added three

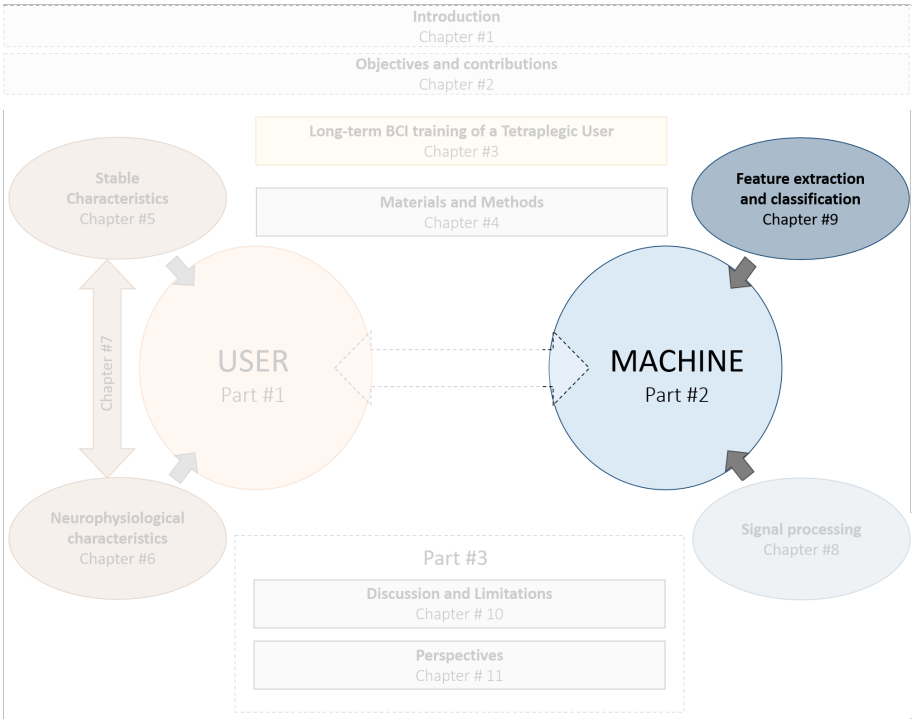
parameters in addition to the *MDFB width* and *mean*. We modeled each correlation signal as a Gaussian function of the frequency and used the parameters of the Gaussian model as predictors (i.e., its mean and variance and the maximum value of the curve).

The prediction models included three factors: The *MDFB width* and the maximum value of the Gaussian model with strong weights compared to other factors, and Gaussian mean value with a smaller negative weight. This result reinforced our previous results, i.e., subjects with a higher *MDFB width* tend to perform better than subjects with a lower one and subjects with an *MDFB mean* value above 16 Hz (in the β band) seem to have lower performances. In addition, we were able to find a new predictor of MI-BCI performances: the maximum value of the optimized Gaussian function but also the corresponding frequency f0. In the final part, we decided to use these coefficients to chose a subject-specific algorithm (constrained or unconstrained) in order to possibly increase online performances. Our results revealed that using a constrained algorithm could help improve online performances of subjects with either distinct EEG signals or no distinct EEG signals and a frequency f0 in the low β band. All together, our studies revealed that studying features extracted from data-driven methods could be interesting to better understand why some subjects have difficulties controlling a BCI. It could also give an indication about the path to follow to adapt and improve these methods by adding relevant constraints. Therefore, future works could consist in exploring similar approaches for other machine learning algorithms used in BCIs, such as spatial filters or classifiers. It could enable us to identify various characteristics of successful spatial filters and classifiers, and possibly improve such algorithms by enforcing these characteristics, if they are causally linked to higher performances.

9

Characteristics of the classifiers

ROADMAP -



QUICK SUMMARY -

In Chapter 8, we have proposed and studied features (or characteristics) extracted from one data-driven method used during the calibration of some BCI systems. We have shown that predictive models using those features were able to predict online MI-BCI performances of unseen subjects. Therefore, it suggested that studying features extracted from data-driven methods could be interesting to better understand why some subjects have difficulties controlling a BCI. In this Chapter, we propose to study and extract predictors from two frequently used data-driven algorithms during the system calibration (i.e., the common spatial pattern (CSP) algorithm and the Linear Discriminant analyses (LDA)). This Chapter is organized as follow:

- We extracted characteristics from the trained CSP and LDA weights.
- We studied the relationships between MI-BCI performances and subject-specific characteristics extracted from the CSP algorithm and LDA.
- We studied whether statistical models can predict and explain MI-BCI user performance across experiments, based on subject-specific characteristics extracted from the machine (i.e., from the results of the calibration algorithms)

Our results suggested that a large majority of our proposed characteristics were strongly correlated to MI-BCI performances.

Contents

9.1 Review of the literature

In the previous Chapter, we have proposed and studied features (or characteristics) extracted from one data-driven method used during the calibration of some BCI systems. We have shown that predictive models using those features were able to predict online MI-BCI performances of unseen users. Therefore, we wanted to go further in the study of data-driven methods and more precisely we studied two frequently used data-driven algorithms during the system calibration (i.e., the common spatial pattern (CSP) algorithm and the Linear Discriminant analyses (LDA)). Indeed, finding predictors from the results obtained with these methods could give more information about the relevance of extracted features (for CSPs) but also the distribution of the classifier weights (i.e., variance, salience among others) prior to the online training which might avoid a difficult learning for some users.

This Section aims at reviewing the current data-driven methods used to design MT-BCIs classifiers and the interpretability of the results obtained using those algorithms. BCI based on MT are, based on change in the oscillatory activity when performing the mental task. This change of power appears in some frequency bands and some brain areas depending on the tasks. Therefore, it is essential to exploit both the spatial and the spectral information. For instance, for an MI-BCI with two tasks, i.e., left- and right-hand motor imagery, the theoretical features to extract would be the average band power in 8-12 Hz and 13-24 Hz from both C3 and C4 channels (see Chapter 6, for more information about neurophysiological activity underlying motor imagery tasks). Unfortunately, such basic design is far for being optimal as for instance, a minimum of 8 channels is a minimum to obtain reasonable performances [200]. Therefore, data-driven methods are essentials to optimize the calibration of a BCI system. However, data-driven methods include very little neurophysiological prior, and rather trust the (potentially noisy) EEG data recorded during the calibration phase. Therefore, it is important to understand algorithms used during the calibration phase but also how the obtained models (e.g., classifiers or spatial filters weights can be interpreted to obtain information about their relevance.

9.1.1 Preprocessing

Most BCI systems use simple spatial filters or temporal filters as preprocessing in order to increase the signal-to-noise ratio of the EEG signals prior to the extraction of features. Indeed, EEG signals are known to be very noisy, as they can be easily affected by the electrical activity of the eyes or the face (e.g., jaw) muscles. Some commonly used preprocessing methods are exposed bellow.

9.1.1.1 Spectral filtering

Appropriate spectral filters, such as band-pass filters or low-pass filters, can be used to reduce the influence of activities that are lying outside of the frequency regions of interest and therefore restrict the analysis to frequency bands in which we know that the neurophysiological signals of interest are. For instance, for BCI based on motor imagery (i.e., sensorimotor rhythms) it is usual to band-pass filter the data in the 8-35 Hz frequency band, as this band contains both the μ and β bands [17]. However, as explained in the previous chapter (i.e., Chapter 8) some algorithms have been developed to select the most discriminant frequency band (MDBF) for each subject individually [18]. Spectral filters can also remove various undesired effects such as power-line inference (50 Hz in France). One filter widely used in preprocessing of EEG is a Butterworth filter, which has an infinite impulse response (IIR) [201]. Spectral filtering can also be achieved using discrete Fourier transform or finite impulse response [201].

9.1.1.2 Spatial filtering

Various spatial filters are used in BCI systems in order to isolate the relevant spatial information embedded in the signals. This is achieved by selecting or weighting the contributions from the different electrodes and by extension the different spatial regions. The two most popular spatial filters used to reduce the background activity are the common average reference (CAR) and the surface Laplacian filters [187]. The CAR filter subtracts the average value of all the electrodes from the electrode of interest while the Laplacian filter subtracts the average of the surrounding channels from the channel of interest. One more advanced spatial filter method, the common spatial pattern (CSP) [40] is widely used in BCI as it has been proved to be very efficient in different BCI applications [17, 202, 203]. More details about the CSP method is given in the following section.

9.1.2 Common spatial pattern

The goal of this method is to design a matrix of spatial filters $W \in \mathbb{R}^{C\times C}$ (C being the number of channels), that projects the signal $x(t) \in \mathbb{R}^C$ in the original sensor space to new time series $x_{CSP}(t) \in \mathbb{R}^C$ in the surrogate sensor space, whose variances are optimal for the discrimination of two populations of EEG related to two different MT. For the rest of the chapter, we will assume that the two MTs are left- and right-hand motor imagery.

Lets note $X \in \mathbb{R}^{C\times T}$ the matrix representing the raw data, where C is the number of channels and T is the number of time samples per channel. Each vector $x(t) \in \mathbb{R}^C (t = 0, .., T)$ represents the EEG signal at a specific time point. For each of the two conditions (i.e., left- (L) and right- (R) hand motor imagery), the estimates of the spatial covariance matrices $\Sigma^{(c)} \in \mathbb{R}^{C\times C}, c \in$L,R of the EEG signal is calculated by averaging over the trials of each task.

The composite spatial covariance $\Sigma \in \mathbb{R}^{C\times C}$ is given as

$$\Sigma = \Sigma^{(L)} + \Sigma^{(R)} \tag{9.1}$$

and can be factored as $\Sigma = U\Lambda U^T$, where U is the matrix of eigenvectors and Λ the diagonal matrix of eigenvalues. Then the CSP analysis is given by the simultaneous diagonalization of the two covariance matrices:

$$\begin{aligned} W^T\Sigma^{(L)}W &= \Lambda^{(L)} \\ W^T\Sigma^{(R)}W &= \Lambda^{(R)} \end{aligned} \tag{9.2}$$

where W is scaled such as $\Lambda^{(L)} + \Lambda^{(R)} = I$, I being the identity matrix. Let $\lambda_j^{(c)}, c \in L, R (j = 1, ...C)$ be the corresponding diagonal element of $\Lambda^{(c)}$. Each of these elements correspond to the variance of the EEG signals after spatial filtering. Since the sum of two corresponding eigenvalues is always one, a large value $\lambda_j^{(c)}, c \in L, R$ indicates that the corresponding spatial filter yields high variance in one condition an low variance in the other. Therefore, sorting eigenvalues (and corresponding eigenvectors) in descending order will give us set of spatial filters that are optimal for discriminating two classes.

The filtered signal is then given by $x_{CSP}(t) = W^T x(t)$, where each column vector $w_j \in \mathbb{R}^C$ of W is called a spatial filter. The term "N pairs of spatial filters" correspond to the N first and N last eigenvectors in W (i.e., the eigenvectors corresponding to the N largest eigenvalues in both conditions). Therefore, for N pairs of filters we have 2N spatial filters.

One these filters obtained, a CSP feature f is obtained computing the log-variance of these spatially filtered signals for each single-trial epoch by $log_{10}(var(X_{csp}))$:

$$f = log(W^T X^T XW) \tag{9.3}$$

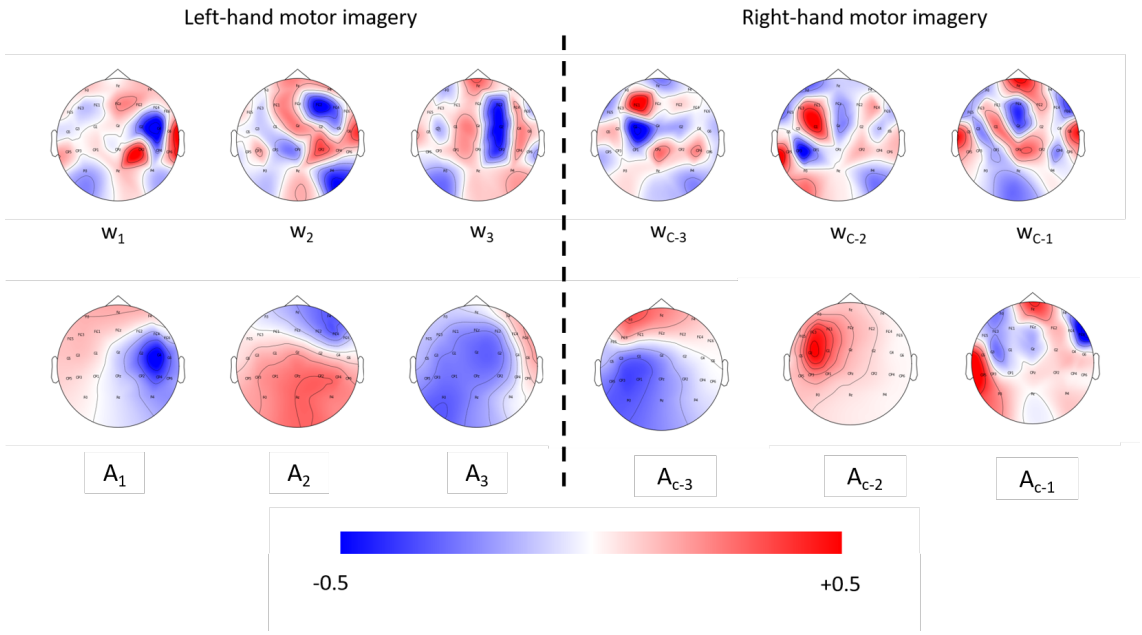


Figure 9.1: Example of CSP analysis for three pairs of spatial filters (6 spatial filters w_j). Their corresponding patterns a_j illustrate how the presumed sources project to the scalp.

9.1.3 Classification

Linear classifiers are discriminant algorithms that use linear functions to distinguish classes. They are probably the most popular algorithm for MT-BCI applications. Two main linear classifiers are used, namely, linear discriminant analyses (LDA) and support vector machines (SVM) [43]. An SVM uses a discriminant hyperplane to identify classes [204]. The selected hyperplane maximizes the margins (i.e., the distance from the nearest training sample). An SVM can be used for linear and non linear classification by using the "kernel trick" which consists in mapping data to other spaces using a kernel function. In our studies, we used LDA classifiers. The objective of LDA is also to use an hyperplane to separate the data representing two classes [40, 205]. To do so, features are transformed into a dimensional space that maximize the ratio of the between-class variance to the within-class variance using the Fisher's criterion. Therefore, LDA is based on linear, i.e., matrix multiplication, transformations. Figure 9.2 gives an example of LDA transformation.

To achieve this goal we need to perform three steps:

- The computation of the separability between different classes (i.e., the distance between the means of different classes): *The between-class variance* S_B.
- The computation of the distance between the mean and the samples of each classes: *The within-class variance* S_W
- The construction of the dimensional space that maximizes the ratio of the between-class variance to the within-class variance:

$$\arg\max_W \frac{W^T S_B W}{W^T S_W W} \tag{9.4}$$

Finally, Equation 9.4 can be reformulated as follow:

$$W = S_W^{-1} S_B \tag{9.5}$$

Finally, the optimal separation of two classes is given by evaluating $w^T f(n) < c$ for a constant c, where $w = \Sigma_\chi^{-1}(m_1 - m_2)$ and where $m_{1/2} \in \mathbb{R}^M$ denotes the classwise means and $\Sigma_\chi \in \mathbb{R}^{M \times M}$

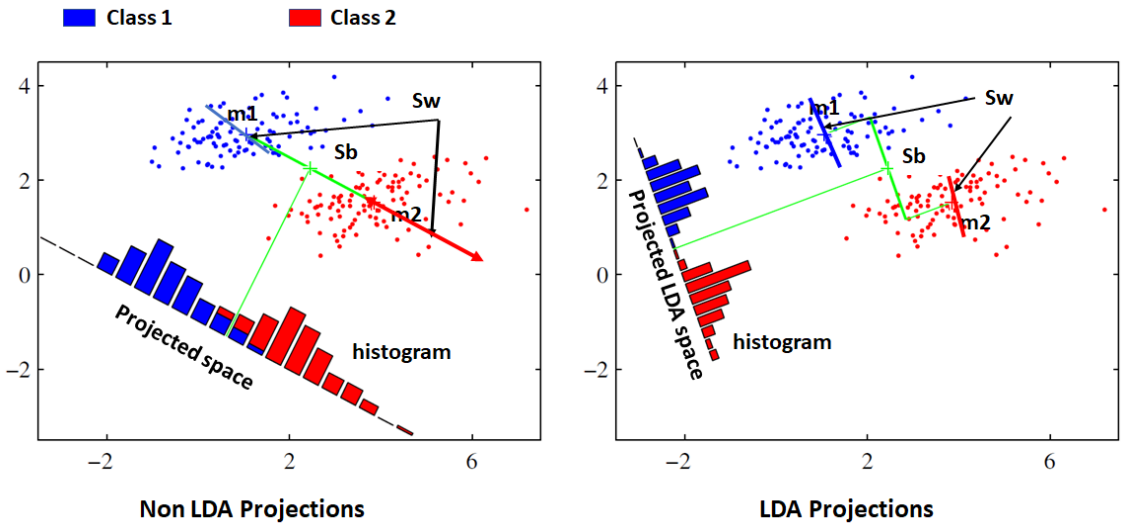


Figure 9.2: Each colour represents a class. m1 and m2 are mean of class1 and class2 respectively. The left plot shows projections of data on the line joining m1 and m2. There is lot of overlaps among samples of two classes. Right plot shows projection of data using LDA hence minimising the overlaps between two classes. S_b represents the between-class variance and S_w the within-class variance. Source: "Pattern Recognition and Machine Learning" by Bishop [206].

represents the common covariance of the observed samples in the feature space. M is the number of features.

9.1.4 Interpretation of weight vectors

Discriminative models or decoding models, extract factors as functions of the observed data. They seek latent factors s_n to approximate known external target variables y_n. These target variables can either be continuous (e.g., reaction times) or discrete (e.g., class label such as left- or right-hand motor imagery). In the linear case, the mapping from observations to factors can be summarized in the transformation matrix $W \in \mathbb{R}^{M \times K}$, with M the number of features (e.g., channels or filters). The decoding model then reads:

$$s(n) = W^T x(n) \tag{9.6}$$

with s(n), the K-dimensional vector of latent factors, W the matrix of filters and x(n) the M-dimensional vector of observed data.

However, the purpose of a filter is to amplify a signal of interest while, at the same time, it should suppress noise. Noise includes technical artifacts (such as head muscles activity) as well as unused features. Therefore, filters W only give information about how to combine information from different features to extract factors (or signal) of interest from data, but not how they are expressed in the features. For instance, CSP filters weights do not give information about how signals of interest are expressed in the measured channels. In the field of neurosciences, if the objective is to interpret filters weights vectors, therefore it is necessary to transform decoding models into encoding models that factorize the data into latent factors $s(n)$ and their corresponding activation patterns A, plus noise $\varepsilon(n)$:

$$x(n) = As(n) + \varepsilon(n) \tag{9.7}$$

For any decoding model the corresponding encoding model is unique and its parameters are obtained by [207]

$$A = \Sigma_\chi W \Sigma_s^{-1} \tag{9.8}$$

Where Σ_χ represents the data covariance matrix and Σ_s the covariance matrix of the latent factors.

To summarize, we can talk about decoding models (how target variables can be decoded from the measurement) as opposed to encoding models (how target variables are encoded in the measurement) [208].

For CSP filters, we call each column vector $a_j \in \mathbb{R}^C, (j = 1, ...C)$ - C is the number of channels- of a matrix $A = (W^{-1})^T$ a spatial pattern (see Figure 9.1 for an illustration of spatial filters and patterns for 3 pairs of CSP) [17, 18]. Applying Equation 9.8 to the LDA filter, the corresponding pattern is $a = m_1 - m_2$

9.2 Objective

In this study, our objective is first to identify potentials predictors of MI-BCI performance based on information extracted from CSP patterns and filters but also LDA patterns and filters. Indeed, as calibration is the first and major step in BCI training, we believe that identifying the relevance of extracted features (for CSPs) but also the distribution of the classifier factors (i.e., variance, salience among others) prior to the online training may help avoid a difficult learning for some users. In addition, it could help to better understand which classifiers and spatial filters characteristics are related to higher control of MI-BCI systems and therefore, in the future, add constraints to optimize classifiers and filters (as we did in Chapter 8).

For instance, in MI BCIs, in theory, relevant spatial information should be extracted around C3 and C4. In addition, we believe that the structure of LDA patterns could also be an indicator of how easy or difficult it will be for the user to control the BCI. We hypothesize that the more there is dispersion in the weights of the LDA pattern, the more difficult the learning will be for the user. Indeed, with no salient weights among the LDA weights (i.e., with all CSP features included with the same importance), within-subject variability might be important as the user might have more difficulties to find the best strategy to properly control the BCI system. Then, to identify new potential predictors and to see if multiple predictors together can improve the performance prediction we created a computational model that could predict the performances of a BCI user based on characteristics extracted from the machine (i.e., from the results of the calibration algorithms).

9.3 Materials and Methods

9.3.1 Datasets

For this study we use the Datasets 2A and 2B where a frequency band selection algorithm was used for online experiments, during the calibration before using the CSP algorithm (see Chapter 8 for more details about this algorithm). Figure 9.3 summarizes both datasets. Three subjects were removed from Dataset 2A due to technical issues during the BCI recording session (e.g., mislabelling of EEG channels that resulted in an erroneous calibration). Therefore, among the 60 available subjects, 57 were included.

9.3.2 Potential factors of prediction

In order to extract different features from CSP and LDA filters and patterns, we used information extracted from OpenViBE, the software used online for computing the CSP filters and the LDA [102]. The six (three pairs) CSP filters were extracted directly using the output provided by the "CSP Spatial Filter Trainer". As an invertible matrix was required to compute the CSP Patterns (see Equation 9.8), we generated 27 (i.e., total number of channels) filters to obtain, with OpenViBE, the general matrix of filters and then computed the patterns associated to the six CSP filters used for the classification. Still using OpenViBE outputs, we were able to obtain LDA filters and patterns. Note that patterns were obtained by using m_1 and m_2 vectors (i.e., classwise means) that were available thanks to a modification inside the source code.

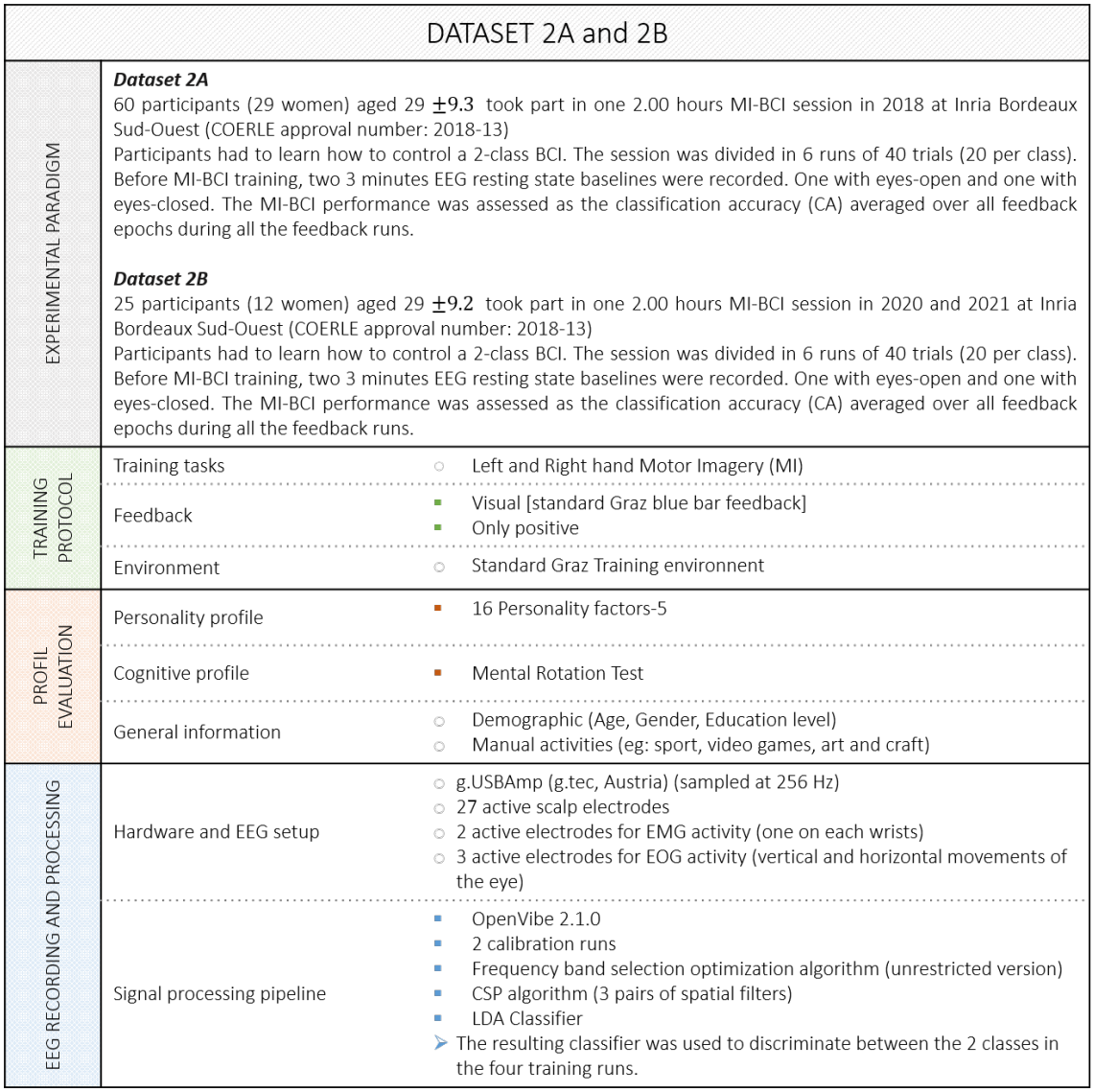


Figure 9.3: Summary table for Datasets 2A and 2B.

In addition, in our case, the first 3 filters (and patters) corresponded to the three largest eigenvalues, i.e., minimal variance for left trials. The first 3 filters (and patters) corresponded to minimal variance for right trials.

For all our analyses, r is the parameter reflecting how far two channels can be to be considered close to each other. In our case we fixed r to the median distance between two channels.

9.3.2.1 From CSP Filters

As explain in Section 9.1, a CSP filter is associated with a weight vector where weights have a complementary role. Indeed, the objective is to extract signals of interest by combining information from different channels. Furthermore, according to the literature, in MI-based BCIs (for left- and right- motor imagery), signals of interest should be located in the sensorimotor cortex and more particularly around C3 and C4 (depending on the task) and surrounding channels should be used to isolate the signal of interest, that should emerge from those electrodes. Figure 9.5 illustrates predictors based on CSP filters. Overall, we extracted the following potential predictors:

– Laterality of filters, to assess if signals that optimally discriminate the two classes are, in majority, located in the right or the left hemisphere. Indeed, left-hand MI should, ideally, activate the right part of motor-cortex and right-hand MI should, ideally, activate the left part of motor-cortex :

$$l_{L/R} = \sum_{C_j \in C_R} \left|w_{C_j}\right| - \sum_{C_i \in C_L} |w_{C_i}| \tag{9.9}$$

with C_L channels of the left hemisphere and C_R channels of the right hemisphere. w_i filter's w weight of channel i.

– The distance D between the two channels corresponding to the two maximal absolute weight of a filter ch_{max1} and ch_{max2} to assess the dispersion of filters:

$$D = \|v_{max1} - v_{max2}\|^2 \tag{9.10}$$

with v_i being the vector containing the 3D coordinates of the channel i.

– Distance D1 between C3 or C4 channels (depending on the filter) and the mean location v of the four channels (with highest weights) surrounding ch_{max1} ($\|v_{ch_{max1}} - v_i\|^2 < \mathrm{r}$) that have a sign opposed to ch_{max1}. Indeed, those four channels (surrounding ch_{max1}) should give information about the localization of the principal source of information (PSI) that best minimize the variance for one specific class. For instance, in filters that minimize the variance for left trials, the mean location of the four channels should be close to C4 channel.

$$v = \begin{cases} x = \frac{x_1+x_2+x_3+x_4}{4} \text{ with } x_i, \text{x coordinate ot the } i^{th} \text{ channel} \\ y = \frac{y_1+y_2+y_3+y_4}{4} \text{ with } y_i, \text{y coordinate ot the } i^{th} \text{ channel} \\ z = \frac{z_1+z_2+z_3+z_4}{4} \text{ with } z_i, \text{y coordinate ot the } i^{th} \text{ channel} \end{cases} \tag{9.11}$$

Then,

$$D1 = \|v_{max1} - v\|^2 \tag{9.12}$$

We then computed, for each predictor P, the mean value of the 6 filters: $L = \frac{1}{6}\sum_{j=1}^{6} |P_j|$

9.3.2.2 From CSP Patterns

CSP patterns show in which channels the signal is reflected. In other words, their weights' value and sign are directly related to the strength and direction of the signal at different channels. Therefore, different features can be extracted from the different CSP patterns:

– The weights variance of each pattern.

– The distance between the channel with the maximal absolute weight and C3 or C4 channels (depending on the pattern): $\|v_{max} - v_{c3orc4}\|^2$ with v_i being the vector containing the 3D coordinates of the i^{th} channels. As users were asked to perform left- and right-hand motor imagery, this feature evaluate how close to C3 and C4 the most informative channel is to see if a sensorimotor area is activated.

– The number of electrodes included in the PSI, i.e, its size. The PSI represents the projection on the scalp of sources that best minimize the variance for one specific class. Channels verifying the following heuristic were included in the PSI:

$$\left|\frac{a_i}{a_{max}} - 1\right| \leq 0.60 \tag{9.13}$$

with a_i being the weight corresponding to the i^{th} channel surrounding the channel with the maximal absolute weight i.e., i for $\|v_{max} - v_i\|^2 < r$.

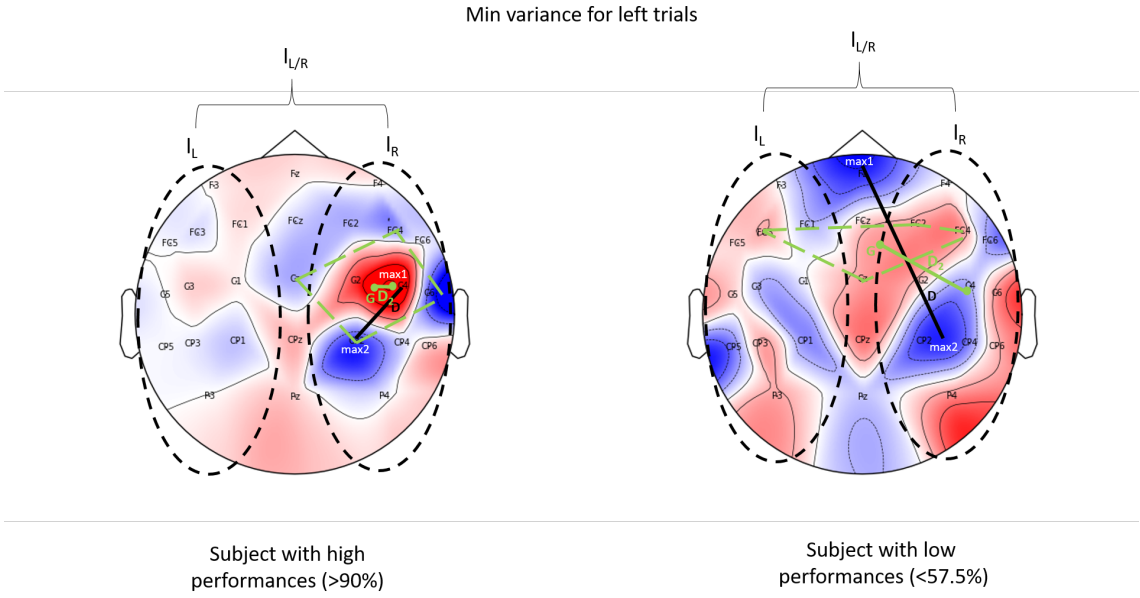


Figure 9.4: Summary of features extracted from the filters. The two filters correspond to the largest eigenvalue of two different subjects. One (left) with high performances and one (right) with low performances. The dotted black circles correspond to right and left hemispheres (and thus, channels included in those hemispheres). The black line corresponds to the distance between the two maximal absolute weight channels. In green, the dotted line represents the four channels (with highest weights) surrounding ch_{max1} that have a sign opposed to ch_{max1}, G is the mean location of those four channels and the solid green line the distance between C4 and G.

– The spatial smoothness P(A) of a spatial pattern A. To estimate it, we used a Laplacian of the form $P(A) = A^T K A$ with $K = D - G$, G being a Gaussian Kernel such that [209, 210]:

$$G_{ij} = exp(-\frac{1}{2}\frac{\|v_i - v_j\|^2}{r^2}) \tag{9.14}$$

$$D_{ii} = \sum_j G_{ij} \tag{9.15}$$

For non-smooth patterns, i.e., patterns in which neighboring channels have very different weights, D, and hence P, will be large.

– Ratio R, representing the ratio between the number of channels with the same sign s and the total number of electrodes. Sign s is the one present in minority in our patterns. Therefore if all the weights of a pattern had the same sign, R was equal to 0. This ratio allow us to determine if a majority of weights have the same sign. This parameter could give additional information about the spatial distribution but also how the power of the main source of information differs from other regions. R was calculated as follow:

w_i corresponds to the weight of channel i.

$Nb_{positive} \leftarrow 0$
$Nb_{negative} \leftarrow 0$
if $w_i > 0$ **then**
  $Nb_{positive} \leftarrow Nb_{positive} + 1$
else
  $Nb_{negative} \leftarrow Nb_{negative}1$
if $Nb_{positive} > Nb_{negative}$ **then**
  $R = \frac{Nb_{negative}}{Totalnumberofelectrodes}$
else
  $R = \frac{Nb_{positive}}{Totalnumberofelectrodes}$

We then computed, for each predictor P, the mean value of the 6 patterns: $L = \frac{1}{6}\sum_{j=1}^{6}|P_j|$

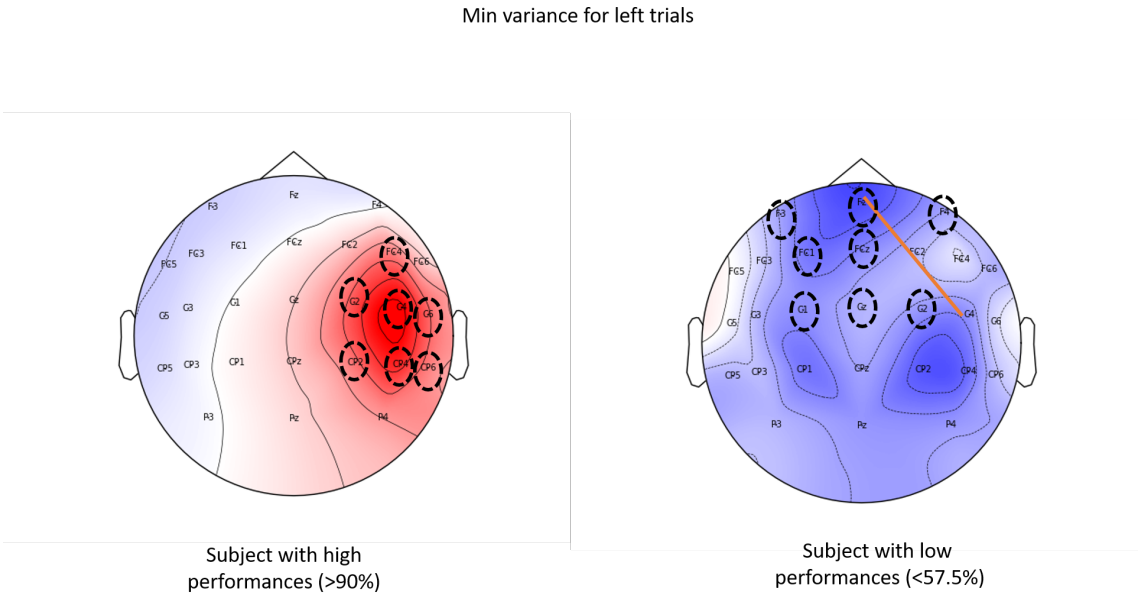


Figure 9.5: Summary of features extracted from the patterns. The two patterns correspond to the largest eigenvalue of two different subjects. One (left) with high performances and one (right) with low performances. Circles in black dotted lines, the number of electrodes included in the principal source of information. The orange solid line corresponds to the distance between the maximal absolute weight channels and C4. For the pattern on the left, $variance = 0.03$, $Spatialsmoothness = 9.42$, $RatioR = 0.40$, distance between v_{max} and $v_{C_{3or4}} = 0$. For the pattern on the right, $variance = 0.006$, $Spatialsmoothness = 2.6$, $RatioR = 0$, distance between v_{max} and $v_{C_{3or4}} = 0.09$

9.3.2.3 From LDA patterns

The objective is to find predictors that can give information about the distribution of the LDA weights.

- Variance V of the LDA patterns' weight.
- Mean of the LDA patterns' weight.
- Mean salience S of the LDA patterns' weight to assess the significance of LDA weights:

$$S = \frac{\sum_i |LDA_i|}{V} \tag{9.16}$$

with LDA_i the i^{th} weight of the LDA.

9.3.3 Analyses

9.3.3.1 Correlations

We performed regression analyses for each predictive factor, to estimate how well they were correlated to the mean MI-BCI performances (i.e., classification accuracy (CA)). As we performed numerous tests, we performed a correction for multiple testing using the Benjamini-Hochberg (BH) procedure [145] to control the false discovery rate (FDR).

9.3.3.2 Models of prediction

In order to build a statistical model to predict/explain MI-BCI performances we used a LASSO regression to obtain models that could predict the performances of MT-BCI users based on subject-specific characteristics extracted from the machine (i.e., from the results of the calibration algorithms), i.e., model 1. Then, we built a statistical model to predict/explain MI-BCI performances based on subject-specific characteristics extracted from the CSP patterns and filters features weighted by the LDA patterns normalized weights, model 2 i.e., $Newf_i = \sum_i w_i \times f_i$, with $Newf_i$ the new weighted feature of the i^{th} CSP pattern or filter and w_i the i^{th} normalized weight of the LDA pattern.

(see Chapter 4, Section 4.6 for more details about LASSO regression and methodology). Table 9.1 summarises the method used to build our models of prediction.

9.4 Results

9.4.1 Outliers detection

We excluded from the analyses the subjects whose mean CA was above or below two SD from the group's mean CA. (see summary in Table 9.2).

9.4.2 Correlations

9.4.2.1 From CSP filters

We conducted correlation analysis between the mean Laterality, the mean distance D, the mean distance D1 and mean online performance (see Figure 9.6) for each group (i.e., Dataset 2A, Dataset 2B and Dataset 2A+2B). After adjustment of the p-values using BH procedure, we found that our predictors were correlated to the mean online performance (see Table 9.3) except for predictor Distance D ($p_{adj} = 0.37$).

In addition, we observed that all predictors were strongly and significantly correlated to each other (see Figure 9.7).

Table 9.1: Summary table explaining the statistical model used to predict/explain MT-BCI performances.

STATISTICAL MODEL OF PREDICTION	
METHOD	LASSO
EXPLANATORY FEATURES	**MODEL 1:** ▪ LDA PATTERNS (MEAN, VARIANCE, MEAN SALIANCE) ▪ CSP FILTERS (MEAN LATERALITY, MEAN DISTANCE D, MEAN DISTANCE D1) ▪ CSP PATTERNS (MEAN VARIANCE, MEAN SMOOTHNESS, MEAN $\|v_{max}-v_{c3orc4}\|^2$, MEAN NUMBER CHANNEL IN PSI, MEAN RATIO R) ➔ 11 features in total **MODEL 2:** ▪ SUM OF CSP FILTERS (LATERALITY, DISTANCE D, DISTANCE D1) WEIGHTED BY LDA PATTERNS NORMALIZED WEIGHTS ▪ SUM OF CSP PATTERNS (VARIANCE, SMOOTHNESS, $\|v_{max}-v_{c3orc4}\|^2$, NUMBER CHANNEL IN PSI, RATIO R) WEIGHTED BY LDA PATTERNS NORMALIZED WEIGHTS ➔ 8 features in total
?	The online trial-wise accuracy, i.e. the default accuracy measure provided online in the MI-BCI scenarios of OpenViBE
EVALUATION OF STABILITY AND RELIABILITY	▪ Leave-one-out cross validations ▪ Random model with 1000 permutations

Table 9.2: Outlier detection: Group details

	Mean group CA	SD group CA	Number outliers	Final Number of subjects
2A	63.32 %	16.39 %	3	54/57
2B	64.90 %	15.13 %	1	24/25
2A + 2B	63.80 %	15.94 %	4	78/82

Table 9.3: Correlation results between each CSP filter factor and online performance for each group. pvalues were adjusted using BH procedure.

	Laterality	Distance D	Distance D1
2A	R= 0.67 ($p=3.10^{-8}$) $P_{adj}= 10^{-7}$	R=-0.45 ($p=7.10^{-4}$) $P_{adj}= 10^{-3}$	R=-0.47 ($p=3.10^{-4}$) $P_{adj}= 6.10^{-4}$
2B	R= 0.60 ($p=2.10^{-3}$) $P_{adj}= 3.10^{-3}$	R= -0.25 (p=0.24) $P_{adj}= 0.24$	R= -0.48 (p=0.01) $P_{adj}= 0.01$
2A + 2B	R= 0.64 ($p=2.10^{-10}$) $P_{adj}= 2.10^{-9}$	R= -0.40 ($p=3.10^{-4}$) $P_{adj}= 5.10^{-4}$	R= -0.48 ($p=8.10^{-6}$) $P_{adj}= 2.10^{-5}$

9.4.2.2 From CSP Patterns

Regression analyses between CSP Patterns mean predictors (i.e., mean of the variance, spatial smoothness, distance between v_{max} and $v_{C_{3or4}}$, number of channels in the principal source and ratio R) and mean online performance (see Figure 9.8) for each group (i.e., Dataset 2A, Dataset 2B and Dataset 2A+2B) revealed after adjustment of the pvalues using BH procedure, that our predictors were correlated to the mean online performance (see Table 9.4) for predictors and all groups except for the mean of the number of channels in the principal source.

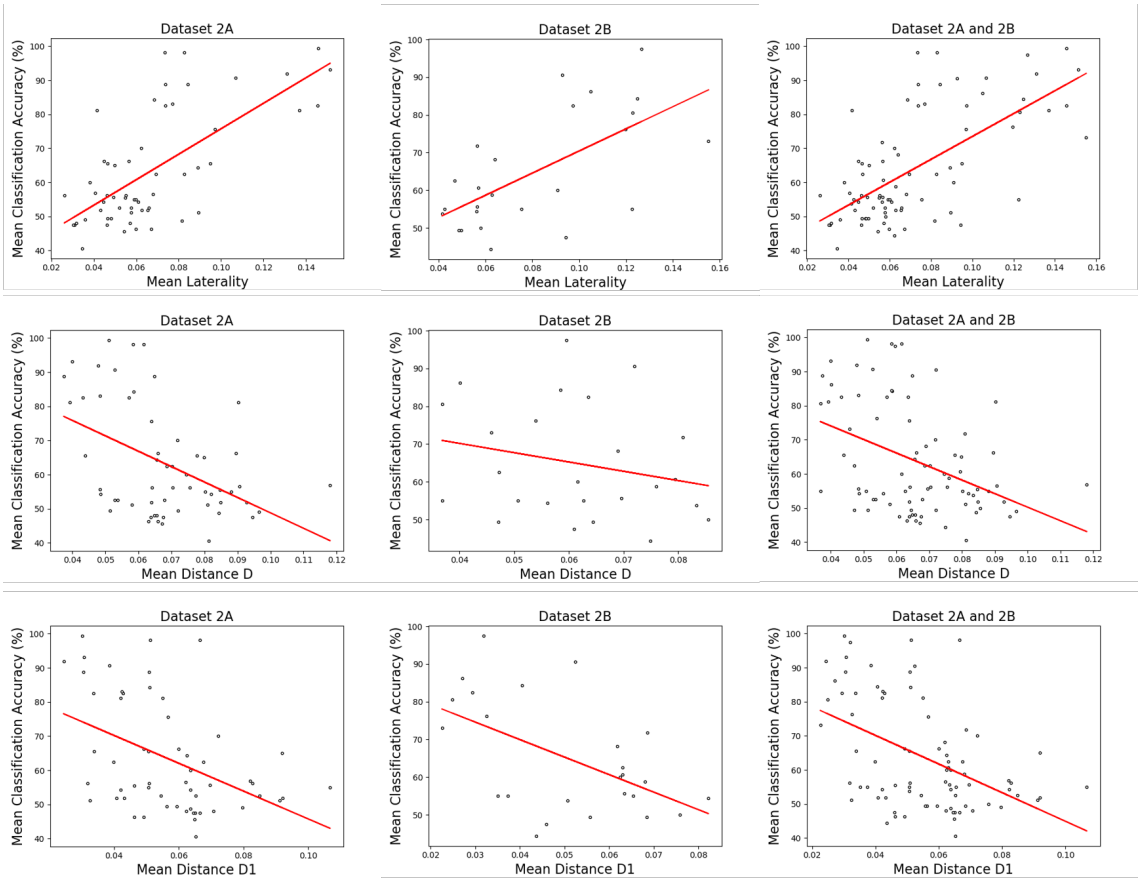


Figure 9.6: Regression analysis between CSP filter factor and online BCI classification accuracy using Dataset 2A, Dataset 2B and Dataset 2A+2B.

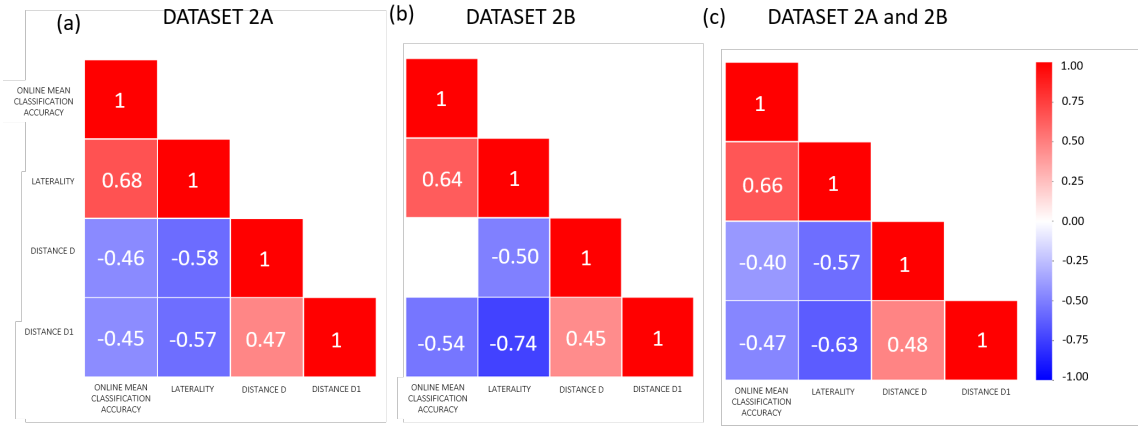


Figure 9.7: Correlation matrices showing correlation score after adjustment of the pvalues (using BH procedure) of Datasets 2A(a), 2B(b), 2A+2B(c). All correlations were significant ($p < 0.05$, at least).

In addition, we observed that most predictors were significantly correlated to each other (see Figure 9.9). For instance, the mean spatial smoothness is correlated to all predictors

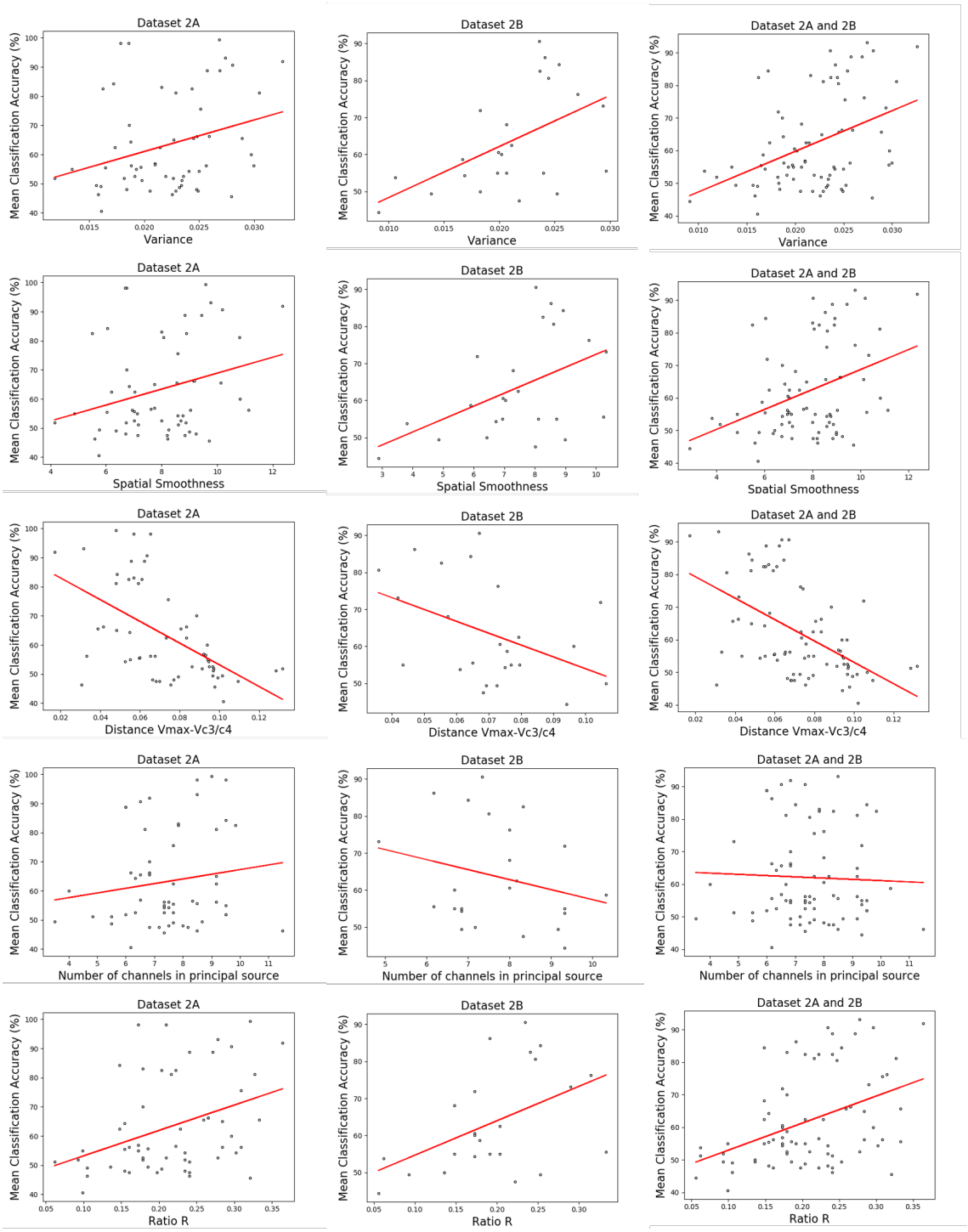


Figure 9.8: Regression analysis between CSP pattern factor and online BCI classification accuracy using Dataset 2A, Dataset 2B and Dataset 2A+2B.

9.4.2.3 From LDA Patterns

Our results revealed strong correlations between LDA Patterns predictors (i.e., Mean, Variance and Mean salience of LDA weights) and mean online performances (i.e., classification accuracy) after adjustment of the pvalues using BH procedure (see Figure 9.10 and Table 9.5).

We observed two outliers with a mean salience above two SD from the group's mean salience. When removing those two subjects, we observed stronger correlations (r=-0.61; $p_{adj} = 10^{-5}$ and r=-0.56; $p_{adj} = 10^{-6}$) for Dataset 2A and 2A+2B.

Table 9.4: Correlation results between each CSP pattern factor and online performance for each group. pvalues were adjusted using BH procedure.

	Variance	Spatial Smoothness	Distance Vmax-$V_{3/4}$	Number of electrodes in the principal source of information	Ratio R
2A	R= 0.30 (p=0.023) P_{adj}= 0.035	R=0.28 (p=0.038) P_{adj}= 0.050	R=-0.56 (p=8.10^{-6}) P_{adj}= 2.10^{-3}	R= 0.15 (p=0.28) P_{adj}= 0.3	R= 0.37 (p=3.10^{-3}) P_{adj}= 8.10^{-3}
2B	R= 0.52 (p=9.10^{-3}) P_{adj}= 0.019	R= 0.47 (p=0.02) P_{adj}= 0.033	R= -0.43 (p=0.04) P_{adj}= 0.050	R= -0.26 (p=0.23) P_{adj}= 0.27	R= 0.47 (p=0.02) P_{adj}= 0.033
2A + 2B	R= 0.42 (p=10^{-4}) P_{adj}= 8.10^{-4}	R= 0.37(p=8.10^{-4}) P_{adj}= 2.10^{-3}	R= -0.53 (p=8.10^{-7}) P_{adj}= 10^{-5}	R= -0.03 (p=0.74) P_{adj}= 0.74	R= 0.41 (p=2.10^{-4}) P_{adj}= 10^{-3}

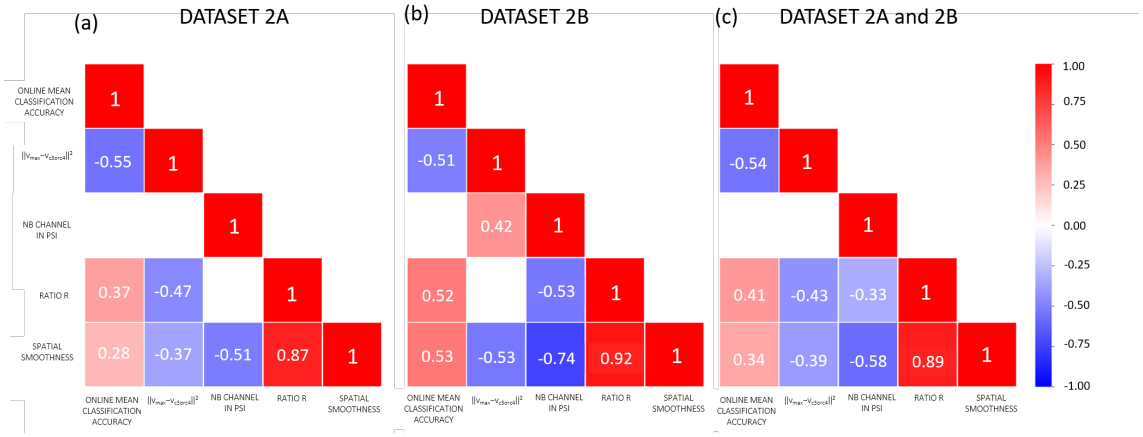


Figure 9.9: Correlation matrices showing correlation score after adjustment of the pvalues (using BH procedure) of Datasets 2A(a), 2B(b), 2A+2B(c). All correlations were significant ($p < 0.05$, at least).

Table 9.5: Correlation results between each LDA pattern factor and online performance for each group. pvalues were adjusted using BH procedure.

	Mean	Variance	Mean salience
2A	R= 0.80 (p=2.10^{-13}) P_{adj}= 10^{-12}	R= 0.70 (p=3.10^{-9}) P_{adj}= 7.10^{-9}	R= -0.44 (p=9.10^{-4}) P_{adj}= 9.10^{-4}
2B	R= 0.67 (p=3.10^{-4}) P_{adj}= 4.10^{-4}	R= 0.63 (p=9.10^{-4}) P_{adj}= 9.10^{-4}	R= -0.66 (p=4.10^{-4}) P_{adj}= 5.10^{-4}
2A + 2B	R= 0.77 (p=2.10^{-16}) P_{adj}= 3.10^{-15}	R= 0.68 (p=9.10^{-12}) P_{adj}= 3.10^{-11}	R= -0.41 (p=2.10^{-4}) P_{adj}= 3.10^{-4}

In addition, we observed that all predictors were strongly and significantly correlated to each other (see Figure 9.11).

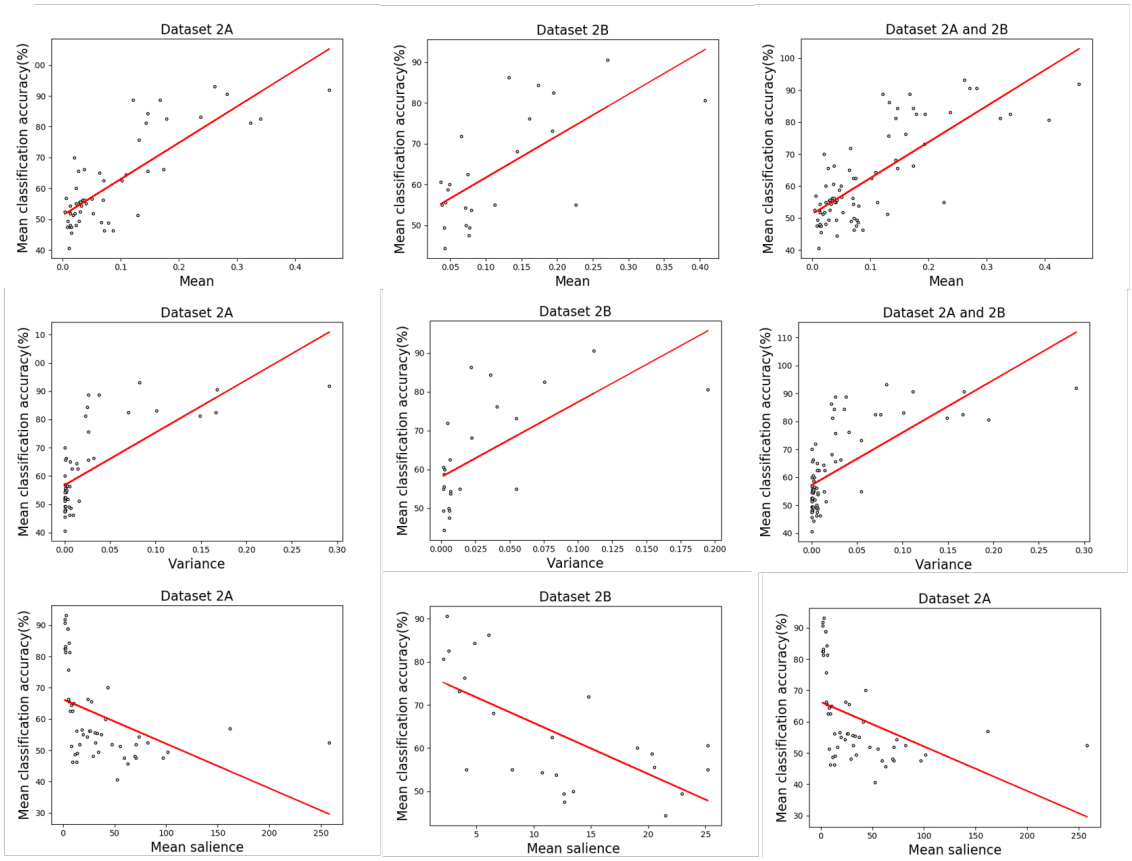


Figure 9.10: Regression analysis between LDA pattern factor and online BCI classification accuracy using Dataset 2A, Dataset 2B and Dataset 2A+2B.

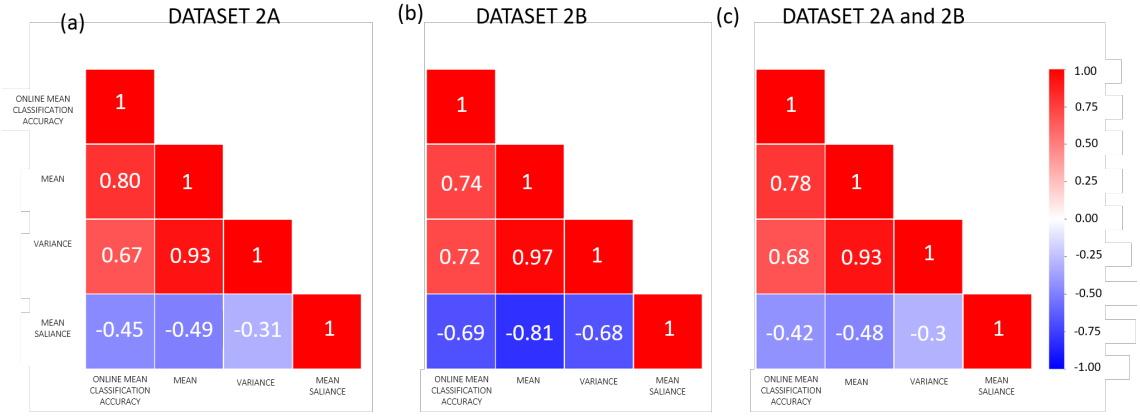


Figure 9.11: Correlation matrices showing correlation score after adjustment of the pvalues (using BH procedure) of Datasets 2A(a), 2B(b), 2A+2B(c). All correlations were significant ($p < 0.05$, at least).

9.4.2.4 Models of prediction

Model 1: Combining all the characteristics extracted from CSP and LDA algorithms results

The results indicated that the predictions made for Dataset 2A were better than chance with a mean absolute error $\sum_{i=1}^{N} \frac{|V_{pred(i)} - V_{real(i)}|}{N}$, n being the total number of models generated, of 7.0%. The mean absolute error (in %) of the random models were 11.02% and 11.47% for $p = .01$ and $p = .05$ respectively (see Chapter 4 for more details about the used method). More precisely, 100% of the generated models included the same three factors (see Figure 9.12):

– The mean value of the LDA pattern weights (Mean LDA), with a strong positive weight of 9.33

– The distance between the channel with the maximal absolute weight and C3 or C4 channels (depending on the pattern) of the CSP patterns ($\|v_{max} - v_{c3orc4}\|^2$) , with a negative weight of -1.4

– The mean ratio R of the CSP patterns which allow to determine if a majority of weights which have the same sign, with a positive weight of 1.5

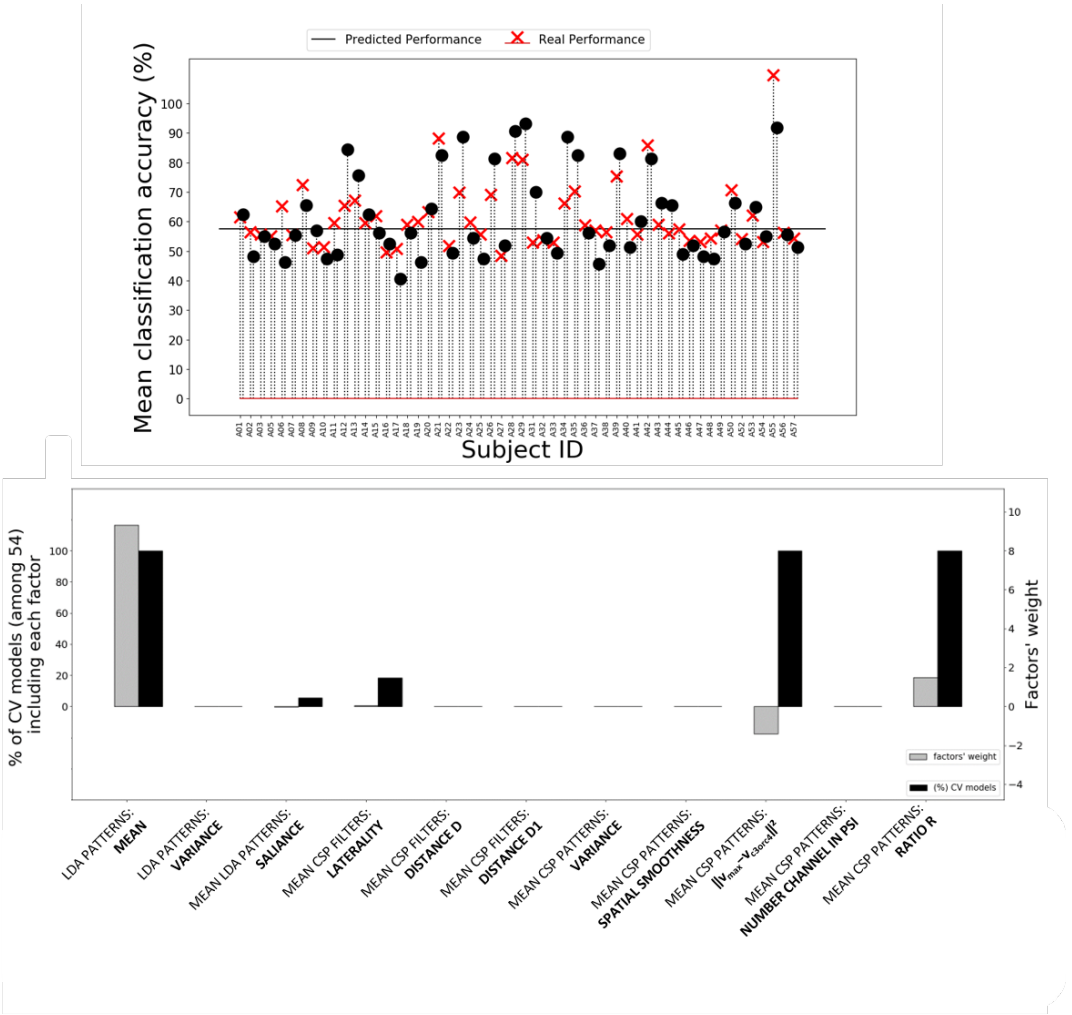


Figure 9.12: Results of the different models (using features for Model 1, see Table 9.1) generated for Dataset 2A. On the top, in black (circle), the real performance of each subject and in red (cross), the predicted performance of each subject generated using the model generated from the training dataset (All subjects except the target one). On the bottom, the percentage of Cross-Validation models including each factor (black) and their weight (grey). The solid black line represents the chance level 57.5% of correct classification accuracy for two classes and 80 trials per class and = 5% [122])

Finally, the final model based on all subjects of Dataset 2A is defined as follow (using normalized factors):

$$Performance(\%) = 61.36 + 9.37 \times MeanLDA - 1.41 \times \|v_{max} - v_{c3orc4}\|^2 + 1.50 \times RatioR \quad (9.17)$$

We computed the predictive performance for each subject of Dataset 2B using Equation 9.17. The mean absolute error $|Perf_{pred} - Perf_{real}|$ was 7.52% (see Figure 9.13) and better than chance as the

mean absolute error (in %) of the random models were 9.99% and 11.13% for $p = .01$ and $p = .05$ respectively. Results showed a positive significant correlation r=0.80 ($p = 2.10^{-6}$) between online performances and predictive performances.

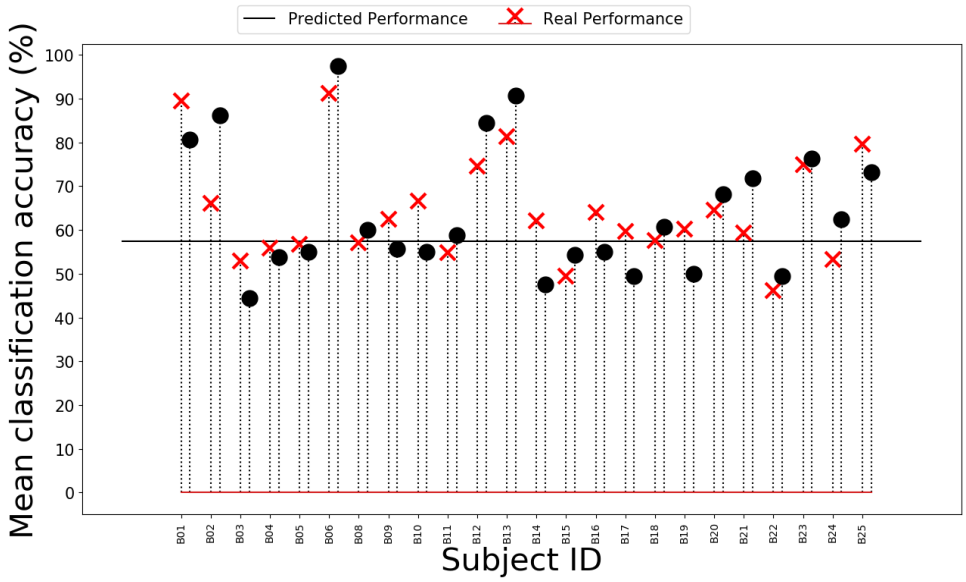


Figure 9.13: Results of the different models (using features for Model 1, see Table 9.1) generated for Dataset 2A. On the top, in black (circle), the real performance of each subject and in red (cross), the predicted performance of each subject generated using the model generated from the training dataset (All subjects except the target one). On the bottom, the percentage of Cross-Validation models including each factor (black) and their weight (grey). The solid black line represents the chance level 57.5% of correct classification accuracy for two classes and 80 trials per class and = 5% [122])

Model 2: Characteristics extracted from CSP filters and patterns weighted by the LDA weights

The results indicated that the predictions made for Dataset 2A were better than chance with a mean absolute error $\sum_{i=1}^{N} \frac{|V_{pred(i)} - V_{real(i)}|}{N}$, n being the total number of models generated, of 7.57%. The mean absolute error (in %) of the random models were 10.79% and 11.52% for $p = .01$ and $p = .05$ respectively (see Chapter 4 for more details about the used method). More precisely, 100% of the generated models included the same four factors (see Figure 9.14):

- The weighted sum of the CSP filters laterality, with a strong positive weight of 6.20
- The weighted sum of the CSP filters Distance D1, with a negative weight of -0.66
- The weighted sum of the CSP patterns Distance $\|v_{max} - v_{c3orc4}\|^2$, with a strong negative weight of -5.06
- The weighted sum of the CSP patterns ratio R, with a positive weight of 0.53

Finally, the final model based on all subjects of Dataset 2A is defined as follow (using normalized factors):

$$Performance(\%) = 61.36 + 6.2 \times WeightedCSPFiltersLaterality - 0.67 \times WeightedCSPFiltersD1 - 5.07 \times WeightedCSPPatterns \|v_{max} - v_{c3orc4}\|^2 + 0.53 \times WeightedCSPPatternsRatioR \quad (9.18)$$

We computed the predictive performance for each subject of Dataset 2B using Equation 9.17. The mean absolute error $|Perf_{pred} - Perf_{real}|$ was 7.57% (see Figure 9.15) and better than chance as the

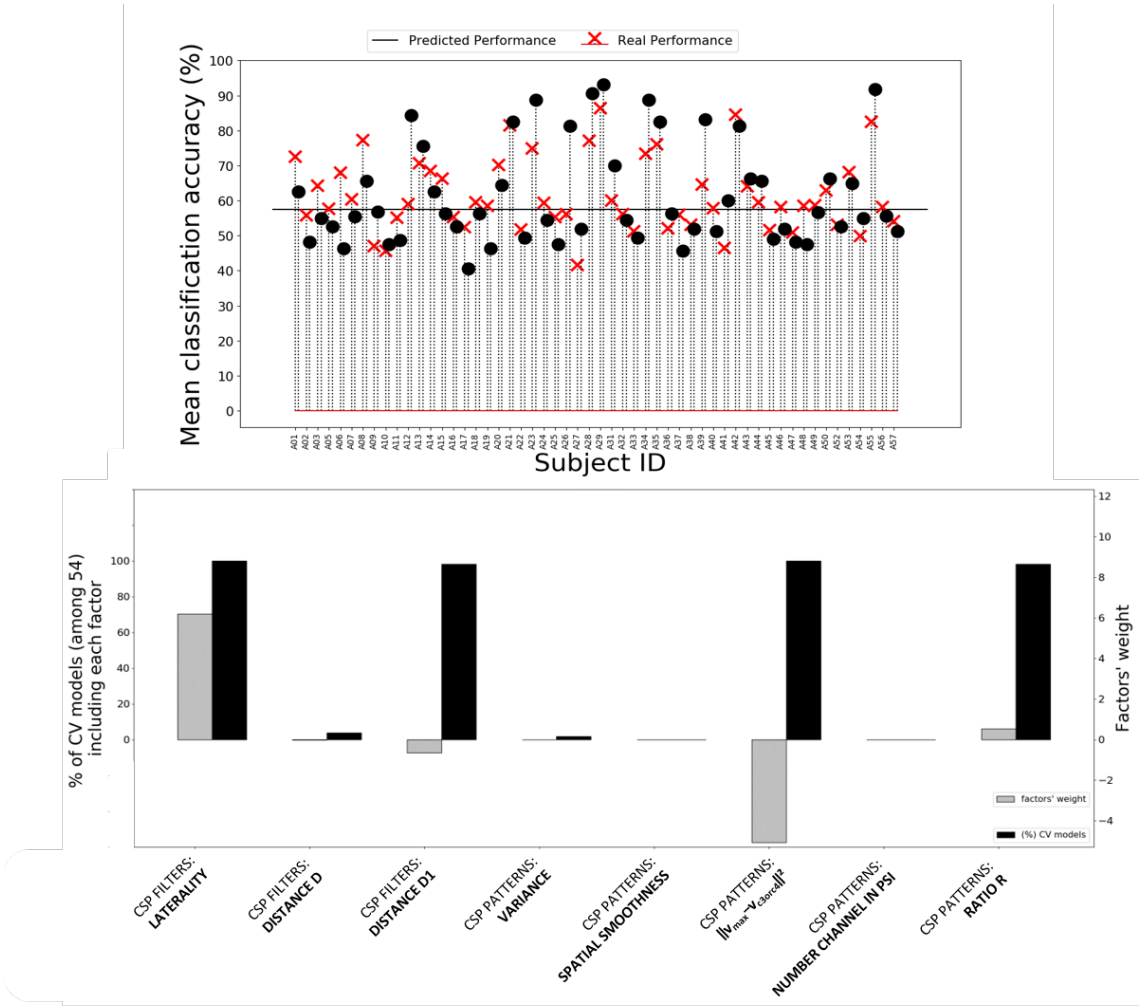


Figure 9.14: Results of the different models (using features for Model 2, see Table 9.1) generated for Dataset 2A. On the top, the percentage of Cross-Validation models including each factor (black) and their weight (grey). On the bottom, in black (circle), the real performance of each subject and in red (cross), the predicted performance of each subject generated using the model generated from the training dataset (All subjects except the target one). The solid black line represents the chance level 57.5% of correct classification accuracy for two classes and 80 trials per class and = 5% [122])

mean absolute error (in %) of the random models were 9.99% and 11.29% for $p = .01$ and $p = .05$ respectively. Results showed a positive significant correlation r=0.80 ($p = 3.10^{-6}$) between online performances and predictive performances.

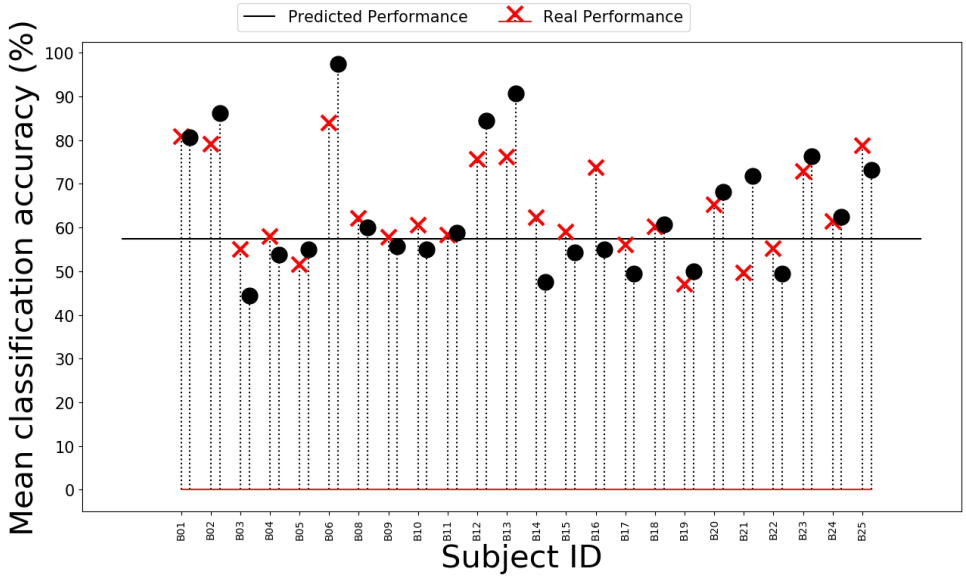


Figure 9.15: Results of the different models ((using features for Model 2, see Table 9.1) generated for Dataset 2A. On the top, the percentage of Cross-Validation models including each factor (black) and their weight (grey). On the bottom, in black (circle), the real performance of each subject and in red (cross), the predicted performance of each subject generated using the model generated from the training dataset (All subjects except the target one). The solid black line represents the chance level 57.5% of correct classification accuracy for two classes and 80 trials per class and = 5% [122])

9.5 Discussion

The CSP and LDA algorithms are successfully used in online MI-BCI systems. One of their advantages is the interpretability of their solutions. Indeed, they are far from being "black-box" methods and we demonstrated that results from those mehods can be interpretaed and used to predict online MI-BCI performances of unseen users.

New factors extracted from CSP and LDA results

In this work we first studied correlation between factors extracted from CSP and LDA results and online performances. Our predictors were all based on neurophysiological expectations when performing left- and right- hand motor imagery. Indeed, when performing those tasks, we expect a lateralization in the brain signals produced by the user. Therefore, the laterality of CSP filters, the distance $\|v_{max} - v_{c3orc4}\|^2$ in CSP patterns (distance between the channel with the maximal absolute weight and C3 or C4 channels depending on the pattern), the distance D between the two channels corresponding to the two maximal absolute weight of a CSP Filter and the distance D1 of CSP filters (localization of the principal source of information (PSI) that best minimize/maximize the variance for one specific class) provide us with information about the expected lateralization. Correlation analyses revealed that all those factors were strongly correlated to each others ($p_{adj} < 10^{-5}$ for all correlations analyses). In addition, we believed that the dispersion of a CSP pattern (or filter), i.e., how distorted a pattern (or filter) can be, could have an impact on online MI-BCI performances. Indeed distorted CSP patterns (or filters) can reveal that they have been affected by outliers. For instance, artifacts, such as blinks and other muscle movements can dominate over EEG signals giving excessive power in some channels. In our factors, the CSP patterns spatial smoothness and variance, the distance D of the CSP filters (distance between the two channels corresponding to the two maximal absolute weight of a filter), the number of channels in the PSI of CSP patterns and the ratio R of CSP patterns (which allow to determine if a majority of weights which have the same sign) give us information about the distortion of patterns and filters. Correlation analyses revealed that, apart from the number of

channels in the PSI of CSP patterns and the distance D of the CSP filters, all those factors were strongly correlated to each others ($p_{adj} < 10^{-3}$ for all correlations analyses)

On the whole, we revealed predictors of MI-BCI performances that represent the lateralization in the brain signals but also the distortion of patterns and filters that could reveal, for instance, artifacts or general noise in the signals.

In addition, factors extracted from the LDA patterns illustrate the distribution and the importance of the LDA weights. Indeed, if all (or a majority) of LDA weights have the same importance for the classification, it might be more difficult for the user to obtain stable online MI-BCI performances and/or to find the optimal strategy, and therefore to learn how to control the BCI system. Indeed, the pattern would be more complex as there would be more features to control. Therefore, among our factors, the LDA patterns variance and salience assess the LDA weights distribution. Furthermore, a majority of LDA weights close to zero (i.e., that do not have a strong impact during the classification), could reveal uninformative features, which are more likely to be noise and therefore could affect the classification.The mean value of LDA patterns but also the salience were used to evaluate this property. Correlation analyses revealed that all factors related to the LDA patterns were strongly correlated to each others ($p_{adj} < 10^{-3}$ for all correlations analyses).

Model of prediction

Another objective of this work was to study the feasibility of predicting/explaining MT-BCI performances across experiments, using a statistical model based on the subject-specific characteristics extracted from the machine. The objective was, through those models of prediction, to identify the stable factors across experiments that may be related/associated to subject performances and thus, on the longer-term, improve BCI user training protocols. In order to have interpretable models we used the LASSO that promotes sparse solutions. We were able to find two models based on one dataset, significantly better than chance and we could prove its reliability by using it on a new set of data of another experiment which had the same training protocol. Factors with the strongest impact in our model were related to the importance of LDA patterns weights (i.e., mean of LDA patterns weights in Model 1), the laterality of CSP filters (CSP filters laterality , CSP filters distance D1 in Model 2), the laterality of CSP patterns (Distance $\|v_{max} - v_{c3orc4}\|^2$ in Models 1 and 2) but also the distortion of CSP patterns (Ratio R in Models 1 and 2). Therefore, if LDA weights do not have a strong impact during the classification, subjects might have difficulties to find a reliable strategy and therefore to control a BCI. In addition, a good lateralization in CSP filters and patterns could also be a good indicator of MI-BCI performances (more precisely for left- hand right-hand motor imagery). Finally, as distortion of CSP patterns might indicate noise in the signal, it could provide guidance on when to give more precise instructions to the user, such as avoiding head muscles activity.

Our results first enhances the importance of checking the plausibility and to investigate neurophysiological properties of data-driven methods. They are preliminary results and the work presented in this Chapter left some questions unanswered and, as such, some future work. Indeed our previous results (Chapter 8 revealed that putting constraint on data-driven methods might have a positive impact on online MI-BCI performances. However, results obtained in Chapter 8 revealed that using a constrained algorithm could help improve online performances of some specific subjects but not all of them. Therefore, it would be interesting to study, in a new experiment, the influence of constraints on both CSP and LDA algorithms to study any potential causal effect of those machine learning factors of BCI performances. Finally, it would be interesting to visualize and extract features from a general feature combining CSP and LDA. Indeed, as showed in Equation 9.3, CSP features f are obtained computing the log-variance of the spatially filtered signals. Therefore, as the optimal separation is given by evaluating $w^T f(n) < c$ for a constant c (using LDA algorithm), we could assume that studying $w^T log(W_{CSP}^T X^T X W_{CSP}) < c$ could also revealed interesting results in terms of performance predictors.

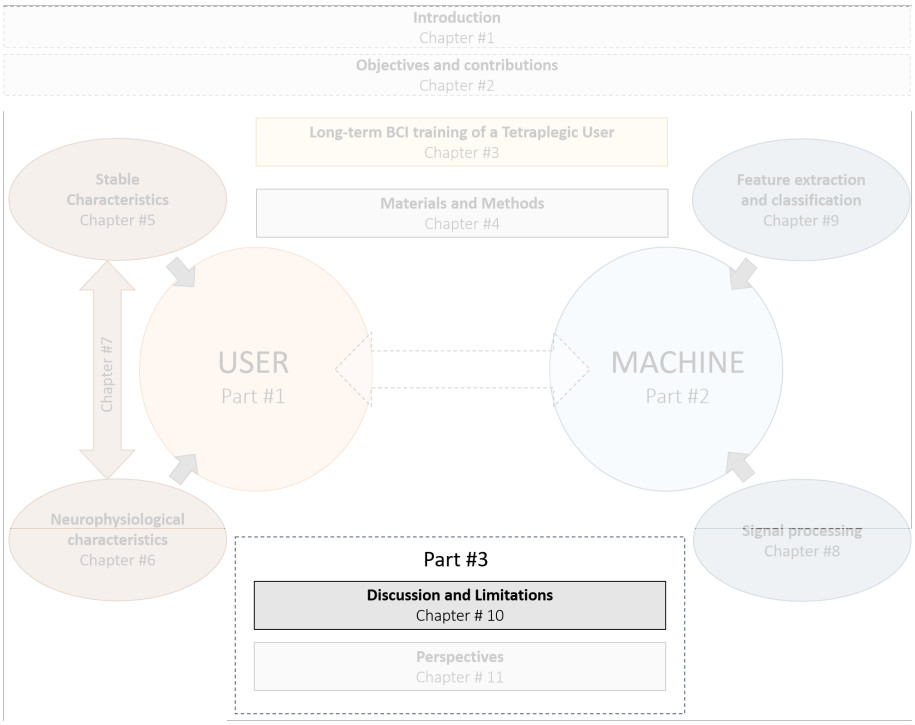
Part III

Discussion and Perspectives

10

Discussion and Limitations

ROADMAP -



To complete this manuscript, in this chapter we will discuss the different contributions and their limitations

10.1 Discussion

Our main objective was to better understand BCI training in order to improve BCI training protocols and thereby BCI efficiency. To best meet this objective, we proceeded in two steps 1) the study of a multi-class MT-BCI longitudinal training (20 sessions over 3 months) of an end-user (i.e., tetraplegic user), 2) the design of predictive computational models of MT BCI performances using both intrinsic user-related factors (i.e., user's traits and neurophysiological characteristics) and extrinsic factors (i.e., characteristics extracted from machine learning algorithms).

10.1.1 Long-term BCI training of a Tetraplegic User

The first contribution, carried out in the context of the CYBATHLON BCI race competition, was devoted to the study of a long-term BCI user training with a BCI-naive tetraplegic user, out-of-the-lab, across multiple sessions. During the user training, we faced different challenges related to EEG non-stationarity, user training itself, but also to the organisation around the competition. For instance, due to the exploratory, and non formal nature of this work but also for reasons of practicality, the user training was performed in different environments (i.e., at home, in the lab) that come with different variability sources. Therefore, after several training sessions i.e., eleven, using common spatial patterns (CSP) and Linear discriminant analysis (LDA) in the signal processing pipeline, we used instead, in order to reduce both within-session variability and shifts between training and test sets, an new adaptive approach based on a Riemannian classifier [95, 96]. This proposed adaptive

approach only models and corrects global EEG distribution shifts within and between sessions, using adaptive rebias Fisher geodesic minimum distance to mean (FgMDM) or between-session baseline projection, respectively. It has the advantage not to rely on task-related EEG changes. This way, we hoped that it would only contribute to stabilize BCI feedback, still encouraging users to maintain effort/consolidating strategies or exploring new ones by enabling him to overly relying on the ongoing machine learning doing all the work, as observed in [120]. Furthermore, we observed that the user learned to adapt his EEG signals to the BCI classifier, instead of increasing his EEG separability between classes (the typically studied type of BCI user learning). This potentially open doors to new ways to assess and study BCI user learning in the future. Indeed, it suggests that in addition to (or instead of) learning how to produce strong and stable signals of interest to control a BCI system, a user might also (or instead) learn how to adapt to the machine. In addition, this observation raises the question of the type of feedback to be used. Indeed, during this study we used an adaptation of the Graz protocol [11, 49] where the feedback is known to be highly uninformative for the user [104, 211], as it does not provide any information about the adaptations the user can make. Therefore, by giving additional information to the user, he might achieve a better EEG separability between classes. Naturally, this result, i.e., the adaptation of the user to the classifier, was obtained with a single subject, and would thus need to be further studied with several other subjects, and other BCI designs, e.g., with non-adaptive classifiers (classifiers that remain stable throughout the training). Finally, the pilot's cognitive load showed an increasing trend throughout the training. We can first hypothesize that the decrease in the load relative to increased expertise was lower than the increase associated with the introduction of more complex and demanding exercises. Another line of interpretation would be to consider the causal factors of the cognitive load related to the environment. For example, stress, emotions and uncertainty can limit working memory capabilities [124] and thus increase cognitive load and impede learning. An acceleration of the training schedule (more sessions per week as the competition was growing nearer) and the approach of the competition might have triggered this type of process. We asked our pilot about his feelings and he revealed that he found it difficult to manage the between-session variability in terms of performances as he felt like he had no control over the system. In addition, some sessions were very exhausting. It might therefore be recommended to carefully assess these elements in the future and to develop methods to support the user and help them to reduce the high load resulting from stress, emotions or uncertainty. Working on the instructions given before and during the task may be relevant [12] as for instance, the user do not yet know what a properly executed mental task should be for a BCI to successfully decode it. This subject is specifically studied in our research team. Finally, this whole study strengthens the importance of being able to predict and/or to identify factors that could predict MT-BCI performances of a user to implement the most optimal subject-specific training protocol. Indeed, the CYBATHLON BCI series 2019 was, for our team, a wonderful opportunity to realise the difficulty of setting up a reliable protocol dedicated to a long term BCI training. Because we did not had enough concrete elements beforehand to design a coherent and efficient training, we had to adapt constantly, which made the training far from being optimal.

10.1.2 User-related characteristics

The second part the work carried out in the framework of this thesis was dedicated to the creation of computational models of BCI user training that could predict users' MT-BCI performances based on intrinsic characteristics (i.e., traits and neurophysiological characteristics) and on extrinsic characteristics (extracted from machine-learning algorithm used for classification). In addition we studied the generalization of those predictive models across different experiments. a Least absolute shrinkage and selection operator (LASSO) [125] regression to obtain interpretable models that could predict variables such as users' performances.

In a first part, we studied user-related characteristics. First, we investigated whether statistical models could predict or/and explain MT-BCI user performance across experiments, based either on (1) users' traits (e.g., personality, gender) and/or (2) information regarding users' previous BCI performances. We used the data from 43 subjects from three different experiments in which participants

had to learn to control a three class MT-BCI. The objective was, through those models of prediction, to identify the stable factors across experiments that could have an impact on BCI users' performances and thus, on the longer-term, improve BCI user training protocols. Our analyses suggested that user traits alone may not be enough to predict MT-BCI performances across experiments, at least with data collected from a small number of sessions (here 3 or less). This was a negative result, as it may suggest that users' traits may not have such a strong contribution to BCI performance variations. As any result, this negative result is nonetheless useful and necessary to better understand and model BCI user training, and notably to identify which factors have an influence and which ones do not [147]. However, it seems possible to predict the performance of a session N+1 using the performance obtained at session N and the score of the vigilance factor. As the vigilance factor measures the tendency to trust others' motives and intentions, it might be interesting to provide deeper technical explanations especially to subjects with a high vigilance score, as they could be skeptical and therefore less engaged in the task. Indeed one might think that emotional and skeptical users would tend to be discouraged if their performances are lower than they expected without understanding why. Finally, we believe that, to reliably predict MT-BCI performance using users' traits, we would need larger samples and/or longer-term experiments. This would enable us to maximally average-out inter-session variability and thereby to obtain a more stable mean performance across sessions that could be better related to users' traits.

To go further in the study of user-related characteristics, we considered that, neurophysiological characteristics of the users, extracted from the recordings of both eyes-closed and eyes-open baselines recorded prior any BCI training, could be relevant predictors of MI-BCI performances. In addition we used those different factors (some from the literature [14, 177, 178] and new ones proposed in this PhD) to create a computational model of BCI performance prediction. For this study, we used data of 80 participants from two different (but similar in terms of protocol) experiments in which participants had to learn to control a two class MI-BCI. Our analyses revealed a model based on one experiment that was significantly better than chance. It was also reliable enough to predict performances of unseen subjects of the other experiment. This model included several factors among which were factors related to motor cortex activation, i.e., the SMR predictor of Blankertz et., al [14], one proposed new predictor representing the variance of the SMR predictor across epochs and the α-band power at rest (with eyes-open and closed). This result is consistent with neurophysiological priors [173]. Indeed, during a motor task, the power decrease (ERD) in α-band is interpreted as a reflection of motor cortex activation [173]. Therefore, the higher the difference between the α-band power at rest and the one during an MI task, the more likely it is to successfully control an MI-BCI. The ability to increase and control the α-band power can be learned with training, and more specifically with a neurofeedback (NF) training [212–214]. Therefore, our results, together with those of the literature, further suggest that an NF training might be beneficial for users who have a low and/or not stable α-band power at rest. For instance, Hwang et., al [212] proposed a training in which participants were instructed to continuously attempt to generate cortical activations around the sensorimotor cortex by imagining left- or right-hand movement using a real-time cortical α-rhythm activity monitoring feedback. The two other factors included in our model were new predictors identified in this thesis: the exponent parameter λ of the $1/f^{\lambda}$ noise spectrum (i.e., its slope) at rest with open-eyes and the θ-band power at rest with eyes-closed. The 1/f noise takes its name from the shape of the spectrum since the EEG power spectral density (PSD) is inversely proportional to frequencies. Higher λ (sharper slope) could indicate an increase of signal to noise ratio and therefore less neural noisy spiking activity [188] compared to lower λ. According to the literature, the increase of spiking activity can be related to aging [189], attention [190] or cognitive speed [191]. Therefore, the parameter λ could reflects the subject's attention process which has already been associated with MT-BCI performances [61, 62, 215]. In addition, the cognitive speed is the ability to identify, discriminate, integrate, make a decision about information, and to respond to visual and verbal information [216]. Thus, being able to evaluate the cognitive speed of a user could be interesting to adapt the user training. Furthermore, our results also suggest that an increase of signal to noise ratio at rest might impact positively MI-BCI performances, which goes in the direction of previous

findings [192] that reported that participants who showed difficulties controlling a BCI had a higher noise than those that could control a BCI. In addition, in our models the θ-band power at rest with eyes-closed is included with a strong negative weight. As predicted λ and the θ-band power at rest with eyes-closed are negatively correlated , we could hypothesize that higher θ-band power represents a higher noise in the signal. This suggestion combined with our previous result on the 1/f parameter λ reinforces the assumption that the signal to noise ratio could be related to BCI performances. In addition, it could supports the hypothesis that the functional relevance of the 1/f power in resting state neural activities substantiates the necessity of isolating the 1/f component from oscillatory activities when studying the functional relevance of spontaneous brain activities as high θ-band power at rest could impact negatively MI-BCI performances.

Finally, we hypothesise that by combining users' traits together with neurophysiological characteristics, we could be able to create a more robust and precise model of prediction for MT-BCI performances. Results revealed that we did not obtained a more reliable model by adding user stable characteristics (i.e., traits).

10.1.3 Machine-related characteristics

The third contribution of this manuscript was dedicated to the investigation of the machine characteristics, i.e., of the characteristics of what the machine has learned, in order to predict and explain MI-BCI performances, still with the aim of improving BCI user training.

Firstly, the objective was to study the relationship between BCI performances and characteristics of the subject-specific most discriminant frequency band (MDFB) selected by a popular heuristic algorithm [18]. For this study, we used data of 80 participants from two different (but similar in term of protocol) experiments in which participants had to learn to control a two class MI-BCI. We first investigated whether there was a relationship between the characteristics of the chosen MDFB and the subjects' online performances (i.e., classification accuracy). Then, we studied the causality of their relationship, if any, and finally if we could find parameters that could be used to refine the algorithm to optimize it further. We used the data from two different experiments to study the impact the MDFB selected by an optimization algorithm during the calibration of an MI-BCI system would have on performances. Our analyses suggested that there was a correlation between the selected user-specific frequency band characteristics, i.e., the mean value and the width of the MDFB selected, and the classification accuracy. To study a possible causality link between them, we added constraints to the algorithm, enforcing characteristics associated to higher performances (based on our first results), and used it in an experiment specifically designed for this study (Dataset B. We then compared the online performances obtained for subjects in this experimental group, Dataset B (using the constrained algorithm) with online performances of matched subjects in a control group Dataset A* from Dataset A (using the unconstrained, i.e., original algorithm). This study did not allow us to conclude on a causality link. Indeed, while average online performances obtained with the constrained algorithm were slightly higher than those obtained with the unconstrained algorithm, this difference did not reach significance. One reason could be that adding constraints could be beneficial for some subjects but detrimental for others. Therefore, in a last part in addition to build computational models to predict the MI-BCI classification accuracy of users using parameters extracted from the MDFB optimization algorithms we tried to build a classification model that could chose the optimal algorithm (constrained or unconstrained) for each subject. We were able to predict the MI-BCI classification accuracy of users using parameters extracted from the MDFB optimization algorithms. To build our models, we added three parameters in addition to the *MDFB width* and *mean* and therefore proposed three additional potential predictors of MI-BCI performances. To identify those predictors we modeled each correlation signal as a Gaussian function of the frequency and used the parameters of the Gaussian model as predictors (i.e., its mean $f0$ and variance b and the maximum value of the curve$a$). The prediction models included three factors: i) the *MDFB width* and ii) the maximum value of the Gaussian model a with strong weights compared to other factors, and iii) Gaussian mean value $f0$, with a smaller negative weight. This result reinforced our previous results, i.e., subjects with a higher *MDFB width* tend to perform better than subjects with a lower one and

subjects with an *MDFB mean* value above 16 Hz (in the β band) seem to have lower performances. In addition, we were able to find a new predictor of MI-BCI performances: the maximum value of the optimized Gaussian function and its mean. In a last part, we decided to use those coefficients to chose a subject-specific algorithm (constrained or unconstrained) in order to possibly increase online performances. Our results revealed that using a constrained algorithm could help improve online performances of subjects with either distinct EEG signals or no distinct EEG signals and a frequency f0 in the low β band. This result emphasises the interest of using predictors of BCI performance in order to choose the optimal machine-learning algorithm for each subject. For instance for subjects with very weak SMR modulations and a discriminant frequency in the low β band, using default values (here our constraints), might be better than letting the algorithm learn from data, as there is nothing clear in the data to learn from. However, for subjects with strong SMR modulations it might be more interesting to let the algorithm learn from data.

In a second step, we studied other machine learning algorithms used in BCIs (i.e., the CSP and LDA algorithms) to identify various characteristics of successful spatial filters and classifiers. For this study, we used data of 82 participants from two different experiments in which participants had to learn to control a two class MI-BCI. Our analyses suggested two models based on one dataset, significantly better than chance and reliable enough to predict performances of unseen subjects of another experiment (which had the same training protocol). The first model i.e., Model 1 included features extracted from CSP filters and patterns and LDA patterns. The second model i.e., Model 2 was based on subject-specific characteristics extracted from the CSP patterns and filters features weighted by the LDA patterns normalized weights. Factors with the strongest impact in our model were related to the importance of LDA patterns weights (i.e., mean of LDA patterns weights in Model 1), the laterality of CSP filters (CSP filters laterality , CSP filters distance D1 in Model 2), the laterality of CSP patterns (Distance $\|v_{max} - v_{c3orc4}\|^2$ in Models 1 and 2) but also the distortion of CSP patterns (Ratio R in Models 1 and 2). As, when performing left- and right- hand MI, a lateralization in the brain signals produced by the user is expected, factors included in our models can give information about the user's ability to produce such lateralized signals. In addition, we observed that having a near-zero mean LDA weight have a strong impact during the classification as subjects might have difficulties to find a reliable strategy and therefore to control a BCI. A next step would be to add constraints to those algorithms according to our results (for instance imposing a laterality in CSP filters) or to find a metric to visualise the main source of information more easily and adapt our training accordingly (for instance, a neurofeedback training could help some subjects achieve better lateralisation of brain signals). All together, our studies revealed that studying features extracted from data-driven methods could be interesting to better understand why some subjects have difficulties controlling a BCI. It could also give an indication about the path to follow in the future to adapt and improve these methods by adding relevant constraints.

10.2 Limitations

We hope this work represents a significant step towards more efficient MT-BCI training procedures and therefore, more reliable MT-BCIs. Nevertheless, some general limitations have to be mentioned. First, we mainly used datasets (except dataset 2B) related to experiments conducted between 2014 and 2018 that had some limitations. For instance, in all datasets, a basic standard signal processing pipeline (CSP + LDA) was used. This is not the most advanced or successful technique currently used, but as the main objective of the experiment was to study user-training, a general pipeline usable in different contexts was used. In addition, a large majority of the participants were students, which limits the possibility of extrapolating these results to the whole population. Therefore, our findings can be considered as relevant for young healthy users (i.e., no neurological, psychiatric disorders or motor impairments). Finally, 3-class experiments used in our analyses included a small number of participants (<20) while our 2-class experiments included only one BCI session. All together, we have not been able to confirm our results in large datasets including multiple BCI sessions. Therefore, one might have doubts regarding the reproducibility of our results and, no matter how promising our results are, future work should investigate this point.

In addition, our results revealed some factors that might be related/associated to performances but do not give information about the possible causation. Indeed, there is nothing explicit in the mathematics of regression that state causal relationships and hence one should not explicitly interpret the slope (strength and direction) in a causal manner. Therefore in this work, our goal was to build evidence for good predictive models instead of proving causation.

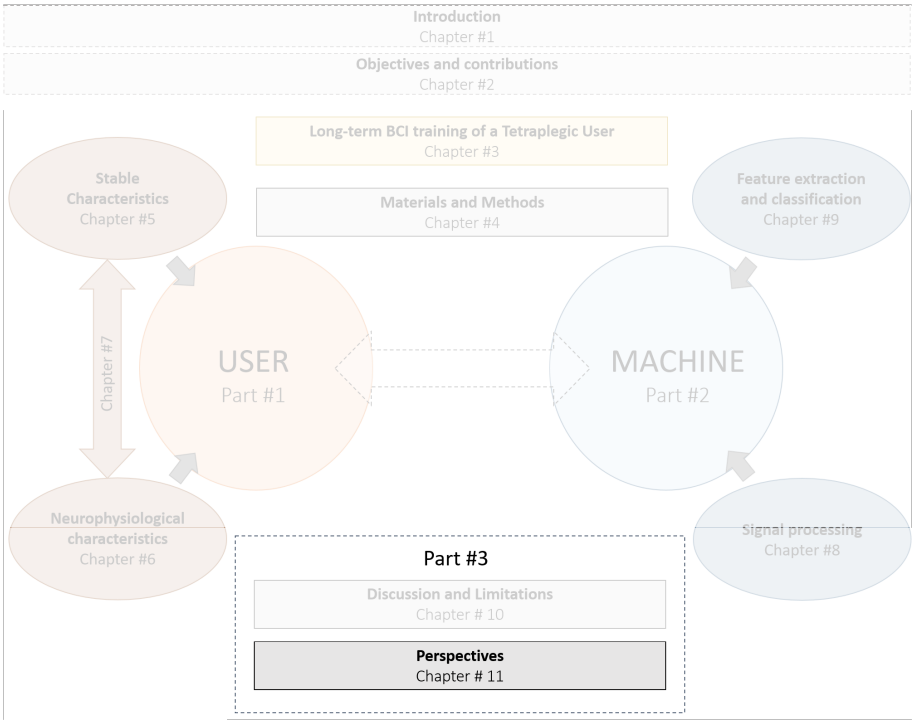
Finally, we believe that it is important to discuss negative results in this PhD manuscript. Indeed, when seeking to publish some of our negative results (Chapter 5), we faced rejections directly linked to the negativity of the results. The given justification was the following: *"Both reviewers agree that the study is relevant, sound and well done. However, the results are negative and therefore the potential impact of this paper will be very low. Authors complain often about the fact that journals do not publish negative results but it is not beneficial from a journal's point of view since there will be few citations if any"*. Scientific research has repeatedly shown that negative results do have a scientific impact and are not only useful, but also necessary for the progress of science. For instance, the BCI scientific community has recently joined forces to co-author an article [217] to stress the need for negative results to be published in BCI research. Indeed, any published scientific result can be a false positive (i.e. a study mistakenly showed that an hypothesis was confirmed) or a false negative (an hypothesis was disproved when it was true) due to chance. In addition, a real negative is interesting, and potentially important, especially for future meta-analysis. Therefore, some published results may be repeatedly considered to be true by the community when it is not, while some genuine results may not be confirmed [218]. This is even more likely to be the case when analyses are conducted on a small sample, as is the case in the BCI community. Furthermore as a large between-subject and within-subject variability is observed in BCI, in terms of brain activity patterns, BCI performances, signal processing method efficiency or in terms of user learning, a result revealed in a given context might not be suitable in another one (results of Chapter 5 are a good example). In conclusion, negative results can enable research teams to save time, allow to confirm or disprove current pieces of knowledge that may be false positives/negatives which is particularly likely in BCI research given the small sample sizes used and the data variability.

11

Perspectives

In addition to future work already mentioned in the previous Section, the PhD thesis work also proposed the way to future long-term research. Some of these aspects are described below.

ROADMAP -



11.1 Different metrics to assess Learning

The principal objective was to better understand BCI training in order to improve the training protocols. To do so, we designed predictive computational models of MT-BCI performances (i.e., classification accuracy (CA)) using both intrinsic user-related factors (i.e., user's traits and neurophysiological characteristics) and extrinsic factors (i.e., characteristics extracted from machine learning algorithms results). However, CA alone, as used in online MT-BCI, is not enough to study BCI user learning and performances [219]. Indeed, for instance, this metric is non-specific, discrete, and directly depends on the classifier and training data. Therefore, it would be interesting to re-do our analyses using different metrics such as the ones presented in Lotte et. al, [219]. Metrics presented in that paper are based on Riemannian distance and were designed to estimate how stable and distinct the EEG patterns for each MI task actually are. Finally, as performance metrics can be complementary using several metrics could enable a better understanding of BCI training. However, doing offline analysis is not enough to study learning. Indeed, the user learns according to the feedback and the feedback depends on the classification methods used. Therefore it is also necessary to collect new data.

11.2 Feedback in the loop

During this PhD, as BCIs are *communication* systems between a *user* and a *machine* we also did a preliminary study dedicated to the study of characteristics extracted from the feedback an the order of the tasks' presentation in order to see the impact of these parameters on users' performances. As feedback is the only information that is provided to a user about how well they can control a BCI system, it seemed essential to study it. During the experimental protocol used in our 2-class datasets, users were asked to imagine the movements of their left and right hands. Based on the classifications made by the system, feedback was provided in the form of a bar the size of which changed according to the user's performance. With regard to MT-BCI feedback, research is often done on the nature of the feedback. For example, the use of feedback in a virtual reality environment [220], feedback based on multiplayer video games, explanatory feedback on a tangible interface [7] and vibrotactile feedback [7, 221]. In addition, researches about biased feedback (only negative, only positive, positive and negative together) during MT-BCI training [222, 223]. However, understanding the impact of feedback characteristics on performance in the context of MT-BCIs remains poor. Our preliminary results suggested that when users' have to perform the same task for several consecutive trials, performances tend to increase suggesting that when the task to be performed is changed too often during a run, users have lower performances. Therefore, it could be interesting to investigate the training program during an online MT-BCI session. The creation of predictive models of MI-BCI performance is still ongoing. We already extracted different characteristics from the feedback related to the feedback size distribution within a trial (e.g., variance, maximal value among others), the time during which the size of the feedback bar is visible (or not) to the user during a trial. In our analyses, we tried to used the features extracted from the feedback from previous trials to predict the performance of the current trial. So far, our results are not conclusive. We are now trying to predict ERD/ERS of a trial instead of the performance.

11.3 The Big Shared MT-BCI Protocol

In this PhD manuscript, we have shown on several occasions that self-regulating one's brain activity through mental tasks to interact is an acquired skill that requires appropriate training and that current training procedures do not enable MT-BCI users to reach adequate levels of performance. So far, most published studies either include a limited number of participants (i.e., 10 to 20 participants) and/or training sessions (i.e., 1 or 2). We still have very little insight into what the MT-BCI learning curve looks like, and into which factors (including both user-related and machine-related factors) influence this learning curve. Finding answers will require a large number of experiments, involving a large number of participants taking part in multiple training sessions. It is not feasible for one research lab or even a small consortium to undertake such experiments alone. Therefore, an unprecedented coordinated effort from the research community is necessary. We are convinced that combining forces will allow us to characterise in detail MT-BCI user learning, and thereby provide a mandatory step toward transferring BCIs "out of the lab". Therefore, during this PhD, an international, interdisciplinary consortium of BCI researchers from more than 20 different labs across Europe and Japan, including pioneers in the field was gathered by Jeunet et. al, [224]. This collaboration will enables us to collect considerable amounts of data (at least 100 participants for 20 training sessions each) and establish a large open database. Based on this precious resource, we could then lead sound analyses to answer the previously mentioned questions. Using this data, our consortium could offer solutions on how to improve MT-BCI training procedures using innovative approaches (e.g., personalisation using intelligent tutoring systems) and technologies (e.g., virtual reality).

Part IV

Appendix

Appendix A: The 16PF5 questionnaire

The Sixteen Personality Factor Questionnaire (16PF) is a self-report personality test developed over several decades of empirical research by Raymond B. Cattell, Maurice Tatsuoka and Herbert Eber. The most recent edition of 16PF, released in 1993, is the fifth edition (16PF5). The test is composed of 185 forced-choice questions in which the participant must choose one of three different alternatives. The test can be done online and last between 30 and 45 min. It provides a score for sixteen primary factors of personality and each factor includes 10 to 15 items.

Once the questionnaire is completed, a raw score is obtained for each factor. Each answer can be worth 0.1 or 2 point. These totals have been created in a way to correlate to the sten scale. Scores on the 16PF5 are presented on a 10-point scale, i.e., a standard-ten scale (sten scale). The sten scale has a mean of 5.5 and a standard deviation of 2, with scores below 4 considered low and scores above 7 considered high. The ranges for sten scores are presented in the Figure 11.1 below. It suggests that there are some limits for the general interpretation and we should not over interpret a potential correlation with a sten score. Indeed, 68% of the population will theoretically obtain a score that falls within plus-or-minus one standard deviation from the mean and the standard error measurement is of 1 sten score point.

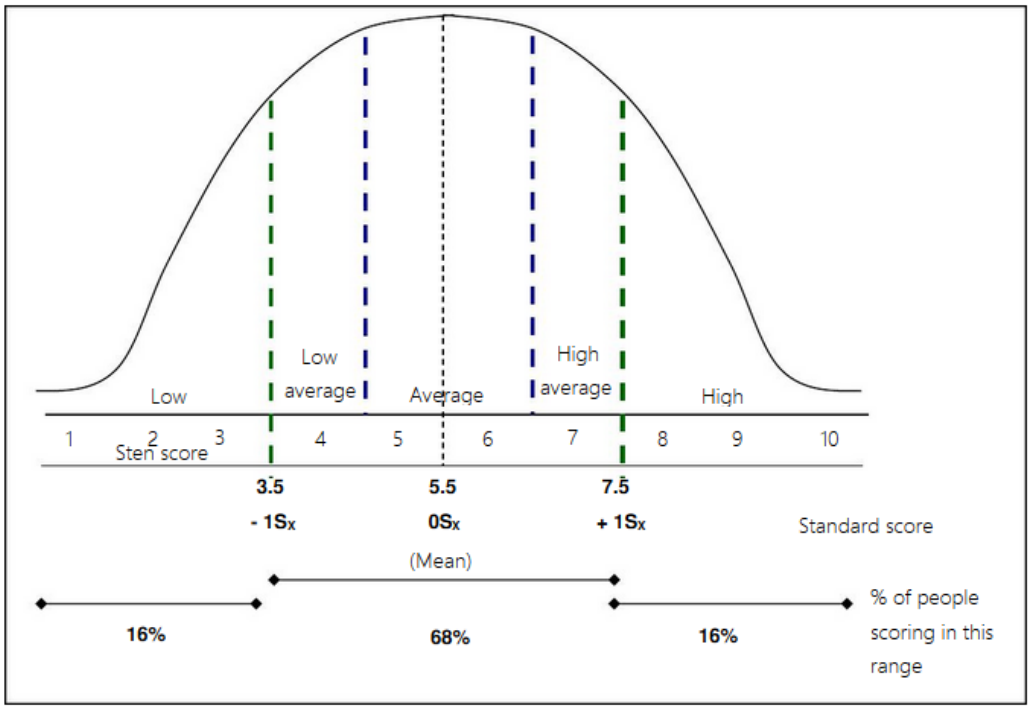


Figure 11.1: Sten distribution (Adapted from Russel Karol 2002).

In addition, the sten scale is a bipolar scale, meaning that each end of the scale has a distinct definition and meaning. Because bipolar scales are designated with “high” or “low” for each factor, a high score should not be considered to reflect a positive personality characteristic and a low score should not be considered to reflect a negative personality characteristic. Table 11.1, is a description of primary and global factors.

16PF5 PRIMARY FACTORS	
WARMTH (A)	Assesses the tendency to get warmly involved in relationships with others (A+) as opposed to the tendency to be more reserved in social and interpersonal relationships (A-).
REASONING (B)	Assesses the ability to solve problems that involve reasoning. High scores (B+) reflect good reasoning skills.
EMOTIONAL STABILITY (C)	Assesses how a person manages his daily life and copes with the difficulties of it. High-scorers (C+) tend to look on the bright side of life and tend to adapt and anticipate more than low-scorers (C-) who feel a certain lack of control over their lives and there is a tendency to react intensely.
DOMINANCE (E)	Assesses the tendency to impose one's will on others (Dominant – E+) or on the contrary to accommodate the desires of others (Respect – E-).
LIVELINESS (F)	Assesses spontaneity. High-scorers (F+) are enthusiastic, spontaneous and seek to be noticed. Low-scorers (F-) are calmer, restrained and considered more mature and sensible.
RULE-CONSCIOUSNESS (G)	Accounts for the intensity with which the social rules that define what is "right" or "wrong" are internalized and used to direct conduct.
SOCIAL BOLDNESS (H)	Assesses the tendency to seek social interaction in a confident, fearless manner and being the center of attention (H+). Low-scorers tend to be shy and timid, and to be more modest and risk-avoidant.
SENSITIVITY (I)	Emphasizes the sensitivity and delicacy of the subjects. High-scorers (I+) tend to base their judgments on personal tastes and ethical values. Low-scorers (I-) tend to have more realistic fields of interest.
VIGILANCE (L)	Assesses the tendency to be confident (L-) or, on the contrary, vigilant (L+) in the face of the motivations and intentions of others.
ABSTRACTEDNESS (M)	Evaluates the type of things people think about and pay attention to. Imaginative subjects (M+) are more oriented towards internal mental processes and ideas than towards aspects of reality. Practical subjects (M-) rely on the data of meaning, observable facts and external realities of their environment to form their perceptions.
PRIVATENESS (N)	Assesses the tendency to be discreet and secret (N+) or, on the contrary, sincere and direct (N-)
APPREHENSION (O)	Assesses the tendency to be anxious or worried (O+) or on the contrary more confident (O-). These feelings are sometimes a response to a particular situation, sometimes a stable characteristic that appears in all situations.
OPENNESS TO CHANGE (Q1)	Assesses the tendency to prefer a rather traditional point of view (Q1-) as opposed to the tendency to think about how improve things and to take advantage of experiences (Q1+).
SELF-RELIANCE (Q2)	Assesses the tendency to seek companionship and enjoy belonging to and functioning in a group (inclusive, cooperative, good follower, willing to compromise) (Q2-). High-scorers tend to be more individualistic and self-reliant and to value their autonomy (Q2+).
PERFECTIONISM (Q3)	Describes a tendency to be self-disciplined, organized, attentive to detail, and goal-oriented (Q3+). Low-scorers (Q3-) have a tendency to leave more things to chance.
TENSION (Q4)	This scale is associated with nervous tension. High-scorers tend to overflow with energy and to become impatient (Q4+). Low-scorers are more relaxed and serene (Q4-). They are patient and not prone to frustration
16PF5 GLOBAL FACTORS	
EXTRAVERSION	Extraverted subjects tend to reach out to others and seek interpersonal relationships. Introverts tend to be less open, sociable, and rather lonely. *EXTRAVERSION = Ax0.3 + Fx0.3 + Hx0.2 + Nx0.3 + Q2x0.3*
ANXIETY	Anxiety can be a reaction to external events or an internal state. High-scorers may find it difficult to control their emotions or reactions and therefore act in counterproductive ways. Low-scorers tend to be unperturbed. *ANXIETY = Cx0.4 + Lx0.3 + Ox0.4 + Q4x0.4*
TOUGH-MINDEDNESS	Intransigent people tend to be reluctant to accept points of view different than their own or to accept new situations. At the other pole, people are warmer and can sometimes lack practical sense and objectivity. *TOUGH-MINDEDNESS = Ax0.2 + Ix0.5 + Mx0.3 + Q1x0.5*
INDEPENDENCE	Independence focuses on the tendency to be strongly and vigorously self-determined, both in thought and in action. *INDEPENDENCE = Ex0.6 + Hx0.3 + Ix0.2 + Q1x0.3*
SELF-CONTROL	Reports on how the subject controls his impulses. High scorers tend to dominate their impulses. *SELF-CONTROL = Fx0.2 + Gx0.4 + Mx0.3 + Q3+0.4*

Table 11.1: Primary and global factors of the 16pf5 personality questionnaire.

Appendix B: Long-term BCI training of a Tetraplegic User

Algorithm 3 TrainFgMDM

Input: $C_{i=1:N_{train}}$: Train SCMs

Output:$\tilde{W_k}$: Fisher Filters, $\overline{C}^{(c)}$:Means of different classes

1: $\overline{C} = argmin_C \sum_{i=1}^{i=N_{train}} d_R^2(C, C_i)$ // (Compute the mean Covariance matrix)

2: $\forall i = 1 : N_{train}, S_i = Log_{\overline{C}}(C_i)$ // (Map SCMs to the tangent space, $T_{\overline{C}}$)

3: $\tilde{W_K} = LDA(\{vec(S_i)\}_{(i=1:N_{train})}$ //vec : vectorize the matrix by stacking the columns of matrix on top of one another // (Compute Fisher Discriminant Filters at $T_{\overline{C}}$, select K first filter)

4: $\forall i = 1 : N_{train}, \tilde{S_i} = \tilde{W_K}(\tilde{W_K}^T \tilde{W_K})^{-1} \tilde{W_K}^T vec(S_i)$ // (Filter all training samples at tangent space, $\overline{T_C}$)

5:$\forall i = 1 : N_{train}, \tilde{C_i} = Exp_{\overline{C}}(unvec(\tilde{S_i}))$ // (Transfer filtered samples on to the manifold)

6: $\overline{C} = argmin_C \sum_{i=1}^{i=N_{train}} d_R^2(C, \tilde{C_i})$ //(Compute the mean of all filtered train samples)

7: $\overline{C}^{(c)}$ = Train MDM classifier($\tilde{C}_{i=1:N_{train}}$) //(Train MDM)

Algorithm 4 Test FgMDM

Input: $C^{'}_{i=1:N_{test}}$: Test SCMs, $\overline{C}^{(c)}$:Means of different classes, $\tilde{W_k}$: Fisher Filters, $\overline{C}$: Mean of train SCMs

Output: $y_i^{'}$:Test label

1: $\forall i = 1 : N_{test}, S_i = Log_{\overline{C}}(C_i)$ // (Map test SCMs to the tangent space, $T_{\overline{C}}$)

2: $\forall i = 1 : N_{test}, S_i^{'} = \tilde{W_K}(\tilde{W_K}^T \tilde{W_K})^{-1} \tilde{W_K}^T vec(S_i^{'})$ // (Filter all test samples at tangent space, $\overline{T_C}$)

3: $\forall i = 1 : N_{test}, C_i^{'} = Exp_{\overline{C}}(unvec(\tilde{S_i}^{'}))$ // (Transfer filtered test samples on to the manifold)

4: $\forall i = 1 : N_{test}, y_i^{'} = argmin_{\overline{C}^{(c)}} d_R(C_i^{'}, \overline{C}^{(c)})$ //(Evaluate test samples by MDM classifier)

Appendix C: Additional correlation analyses

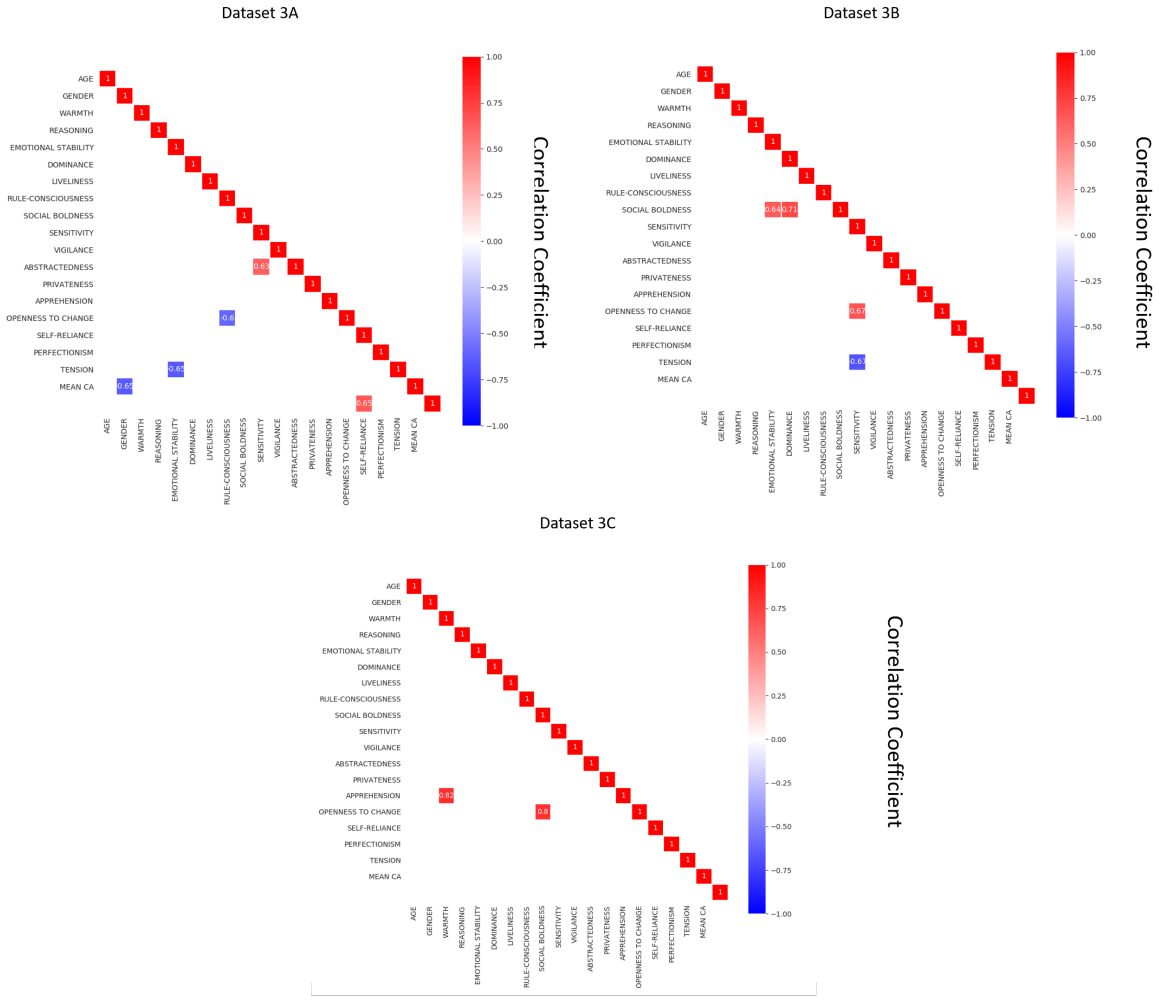


Figure 11.2: User traits Charactristics - Correlation matrix of features extracted from Dataset 3A, 3B and 3C,. Correlation coefficient is reported for correlations with $p_{adj} < 0.01$.

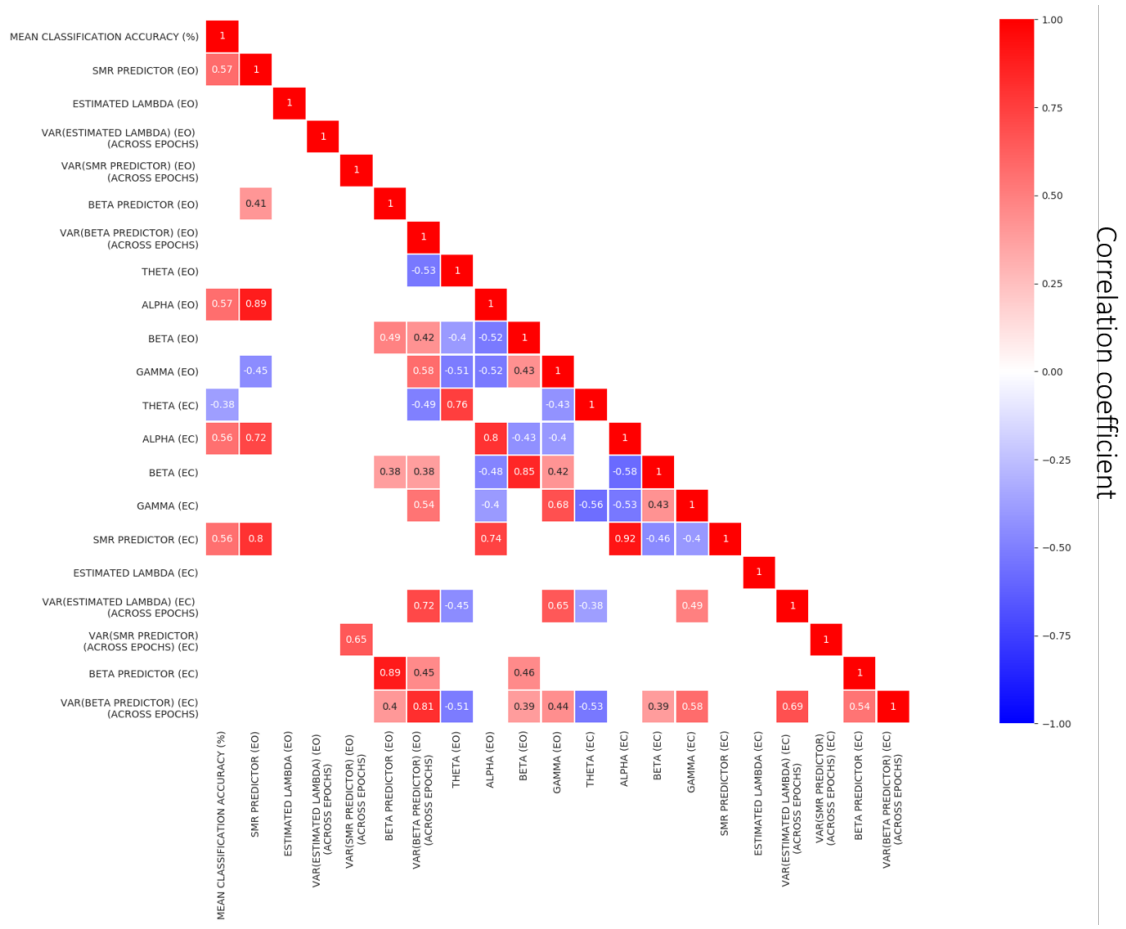


Figure 11.3: Neurophysiological Charactristics - Correlation matrix of features extracted from Dataset 2A. Correlation coefficient is reported for correlations with $p_{adj} < 0.01$.

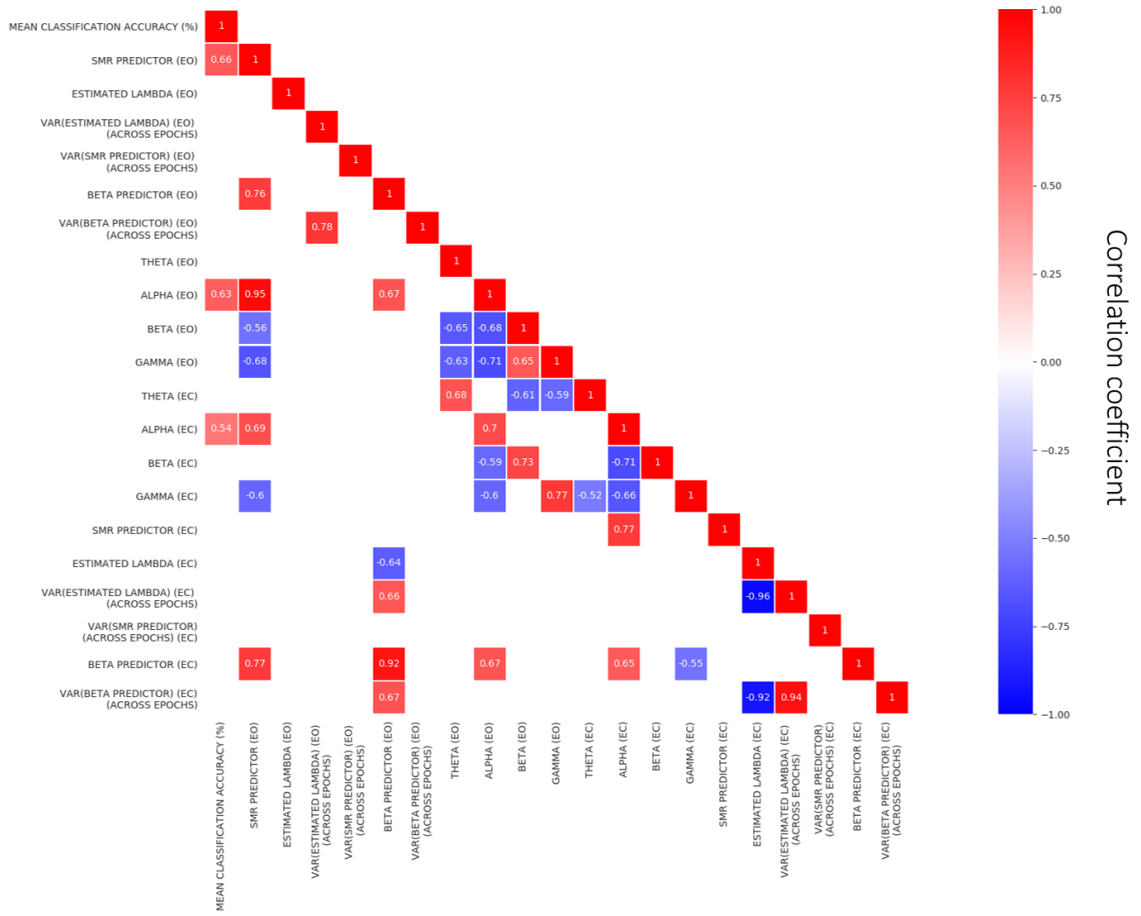


Figure 11.4: Neurophysiological Charactristics -Correlation matrix of features extracted from Dataset 2B. Correlation coefficient is reported for correlations with $p_{adj} < 0.01$.

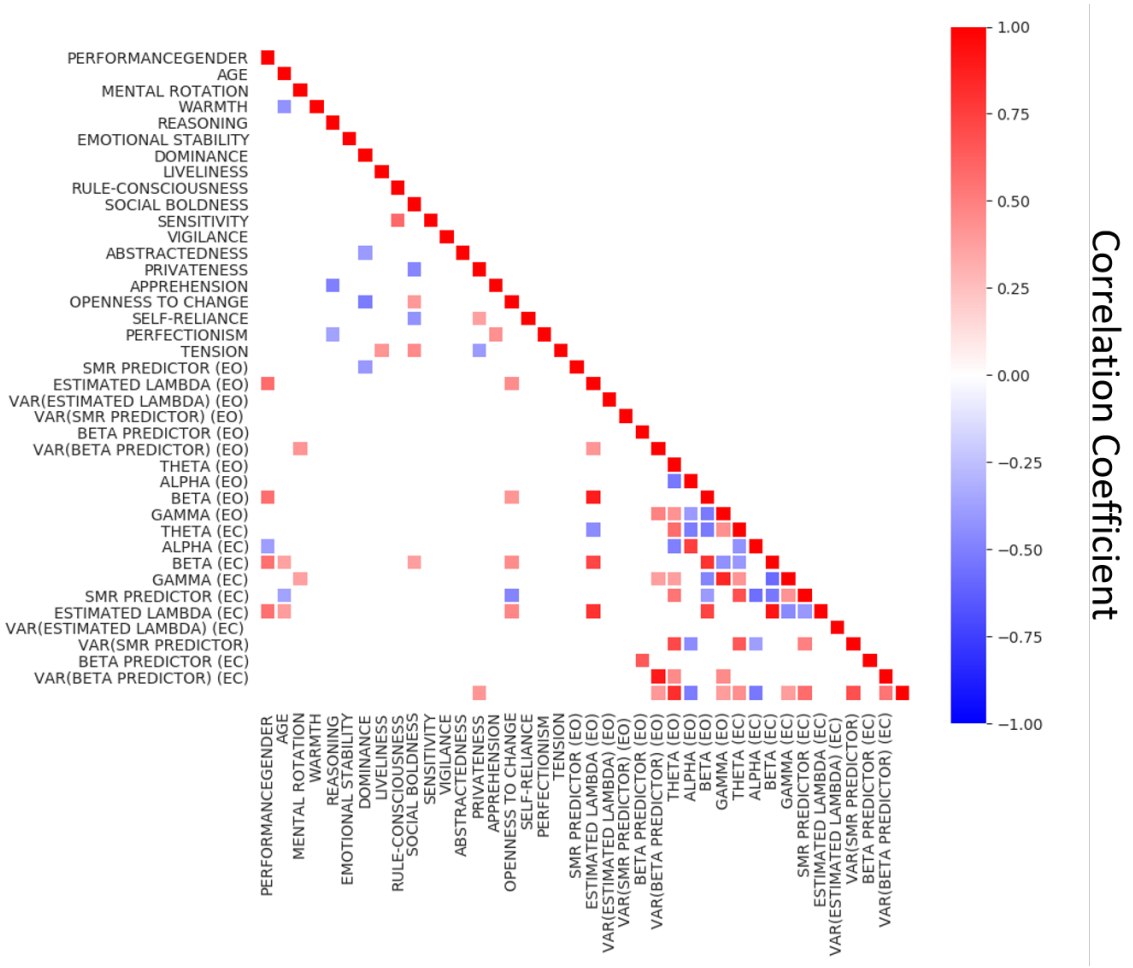


Figure 11.5: Neurophysiological Charactristics and traits -Correlation matrix of features extracted from Dataset 2A. Correlation coefficient is reported for correlations with $p_{adj} < 0.01$.

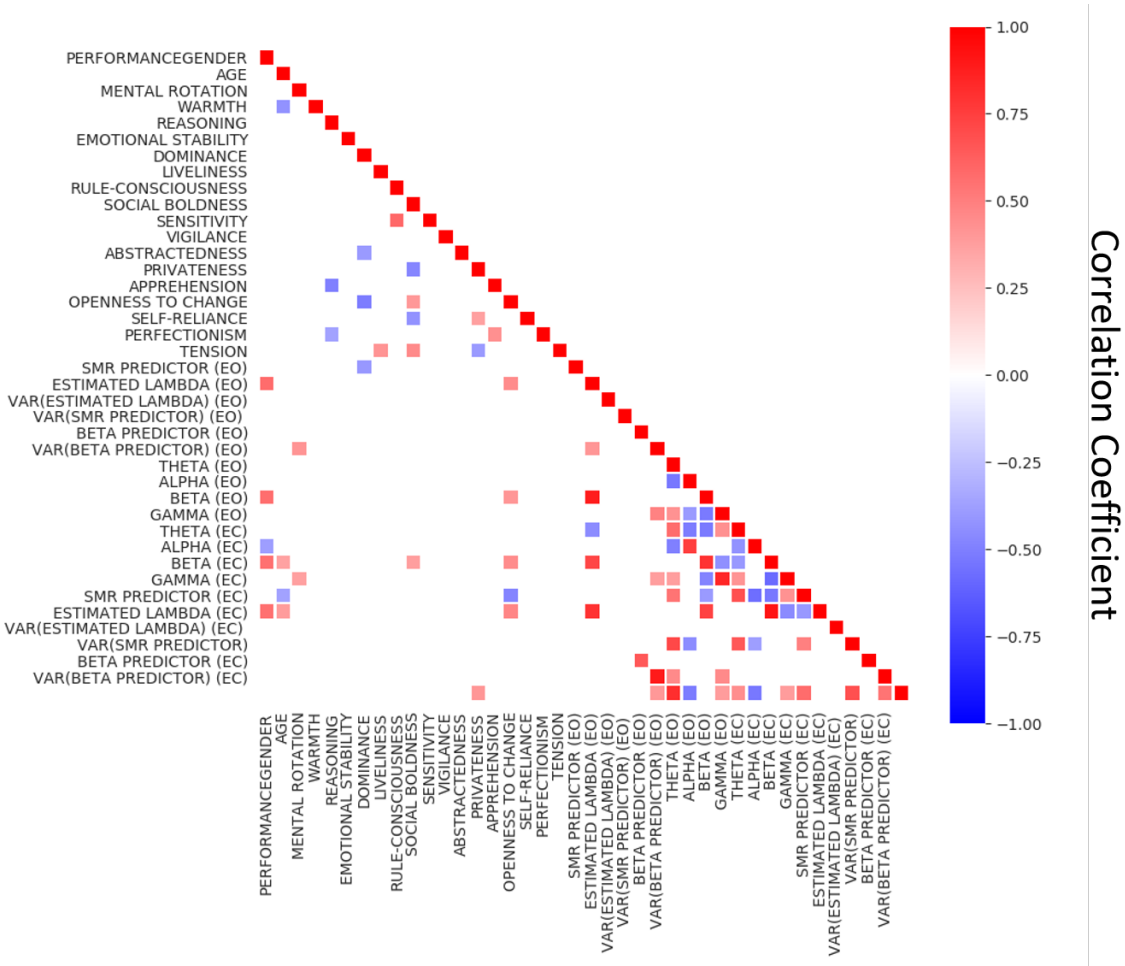


Figure 11.6: Neurophysiological Charactristics and traits -Correlation matrix of features extracted from Dataset 2B. Correlation coefficient is reported for correlations with $p_{adj} < 0.01$.

Appendix D: Elastic-Net Regression

In addition to the LASSO regression used in Parts I and II, we used the Elastic Net (EN) [141] regression that lead to a unique model when it is not the case with the LASSO. Elastic Net is also more robust and stable than Lasso [225] since with Lasso variable selection can be too dependent on data and thus unstable. In addition, EN can handle correlated variables better than Lasso. Indeed, with such variable, Lasso usually randomly chooses one and ignores the others, i.e., one of the coefficients of the correlated features will be non-zero while the others will be zeros. In contrast, Elastic Net will pick both of them with similar coefficient values, thus increasing the number of variables in the model. Therefore, it does not enable a variable selection as sparse and therefore as interpretable as with the LASSO.

Elastic-Net

Elastic Net regression uses $l1$ and $l2$-norm regularization with two penalty parameters λ_1 and λ_2 [225]. This combination allows us to create a sparse model with some weights that are non-zero like Lasso while still maintaining the regularization properties of Ridge regression [225]. To set up the model we need: $y \in \mathbb{R}^n$ which here is the MI-BCI performance vector, a matrix $X \in \mathbb{R}^{n \times p}$ with p features (here the neurophysiological predictors) for n subjects and a coefficient vector $\beta \in \mathbb{R}^n$ which is the regression weight. With Elastic Net, the regression weights are estimated by:

$$\hat{\beta}_{e-net} = \underset{\beta \in \mathbb{R}^p}{argmax} \|y - X\beta\|_2^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2 \tag{11.1}$$

where, $\|u\|_2^2 = \sum_{i=1}^{n} u^i$ for $u \in \mathbb{R}^n$, $\|\beta\|_1 = \sum_{j=1}^{p} |\beta_j|$ and $\|\beta\|_2^2 = \sum_{j=1}^{p} \beta_j^2$. In order to determine the ratio between λ_1 and λ_2, we performed inner cross-validation (leave-one-subject-out cross-validation) on the N-2 subjects (our 'inner' training set) to estimate the optimal ratio, i.e., the one that minimizes the mean absolute error on the training set. Then we used this optimal ratio to build the model for the N-1 training subjects. The left out subject is the test set. This process was repeated by using separately each subject as the test set.

We used the same method as for the LASSO regression to compute empirical chance level for our data. By rearranging the training set BCI performances randomly and keeping the features the same, we broke the before-known relationship between features and performance. We kept the same EN parameters (λ_1 and λ_2) found in the inner cross-validation. Moreover, we kept the same features value and performance score for the test set that was left out to compare the actual model to the random one. We ran the EN model for the same test subject with the randomly shuffled training set. We did this process 10,000 times and stored the mean absolute error to obtain the distribution of the chance prediction performances. We then sorted the values in descending order and computed the 95^{th} and 99^{th} percentiles, which represent the chance level for the mean absolute error for $p = 0.05$ and $p = 0.01$ respectively.

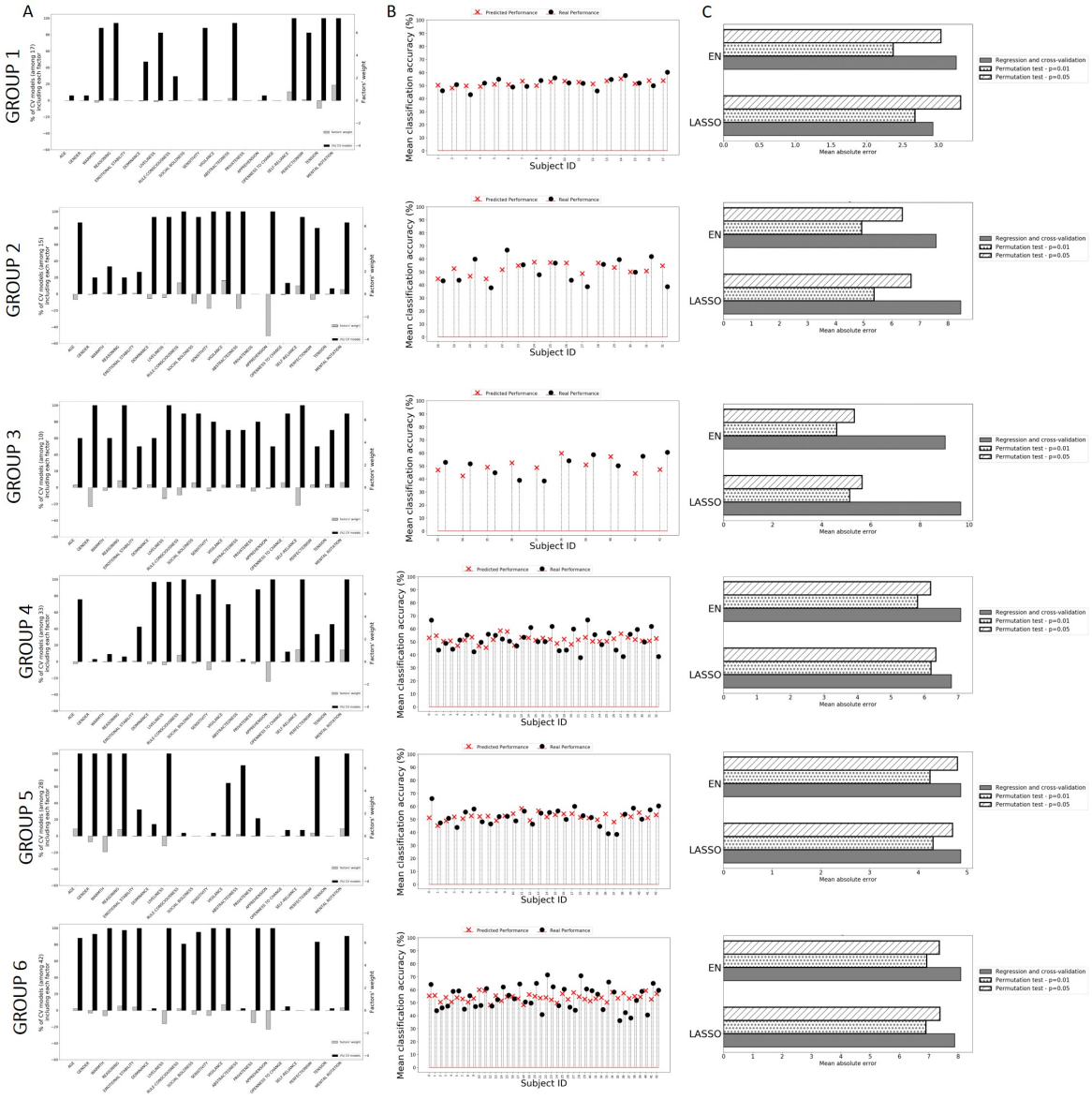


Figure 11.7: Results of the different between-subjects models generated using an Elastic Net regression for each Groups of Chapter 5. A, the percentage of Cross-Validation models including each factor (black) and their weight (grey). B, in black (circle), the real performance of each subject and in red (cross), the predicted performance of each subject generated using the model generated from the training dataset (All subjects except the target one). C, comparison of the results obtained using LASSO and Elastic-NET regressions.

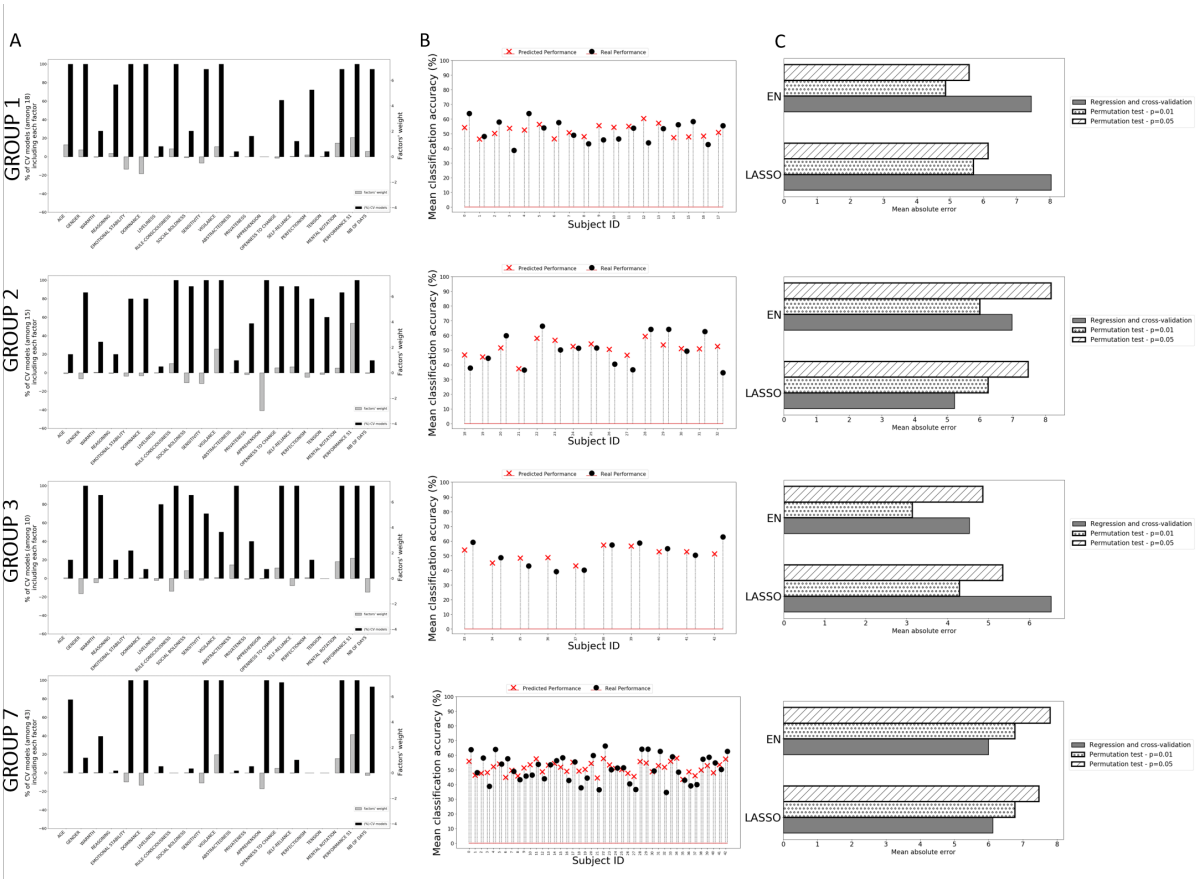


Figure 11.8: Results of the different within-subjects models generated using an Elastic Net regression for each Groups of Chapter 5. A, the percentage of Cross-Validation models including each factor (black) and their weight (grey). B, in black (circle), the real performance of each subject and in red (cross), the predicted performance of each subject generated using the model generated from the training dataset (All subjects except the target one). C, comparison of the results obtained using LASSO and Elastic-NET regressions.

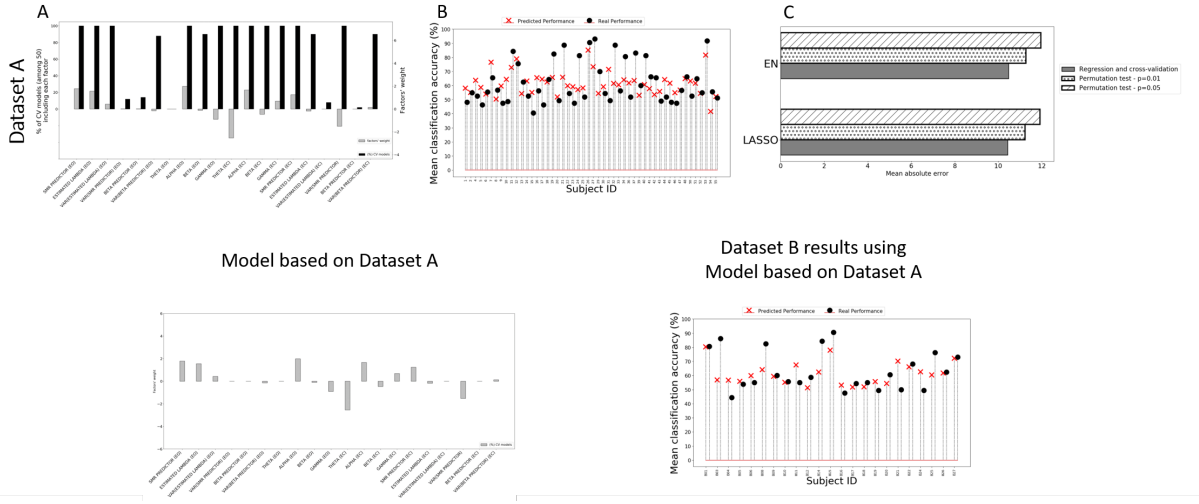


Figure 11.9: Results of the different models generated using an Elastic Net regression and neuropshysological characteristics (see Chapter 6). A, the percentage of Cross-Validation models including each factor (black) and their weight (grey). B, in black (circle), the real performance of each subject and in red (cross), the predicted performance of each subject generated using the model generated from the training dataset (All subjects except the target one). C, comparison of the results obtained using LASSO and Elastic-NET regressions. On botton, the model generated using all subjects from dataset A and the results obtained using this model on dataset B.

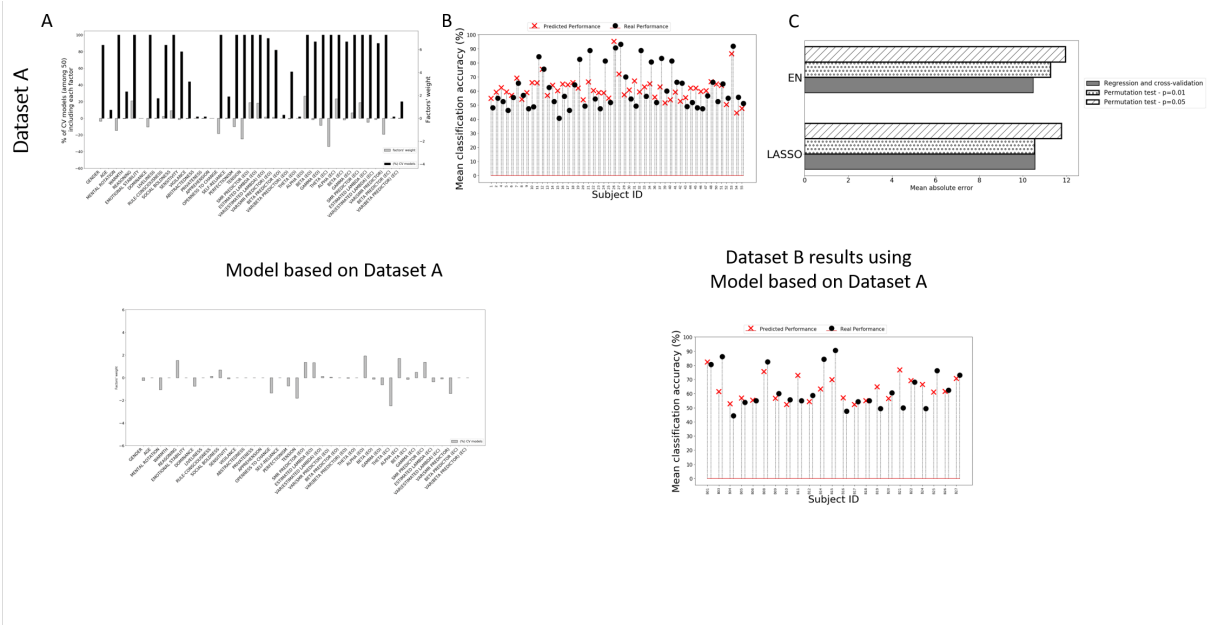


Figure 11.10: Results of the different models generated using an Elastic Net regression and all user-related characteristics (see Chapter 7). A, the percentage of Cross-Validation models including each factor (black) and their weight (grey). B, in black (circle), the real performance of each subject and in red (cross), the predicted performance of each subject generated using the model generated from the training dataset (All subjects except the target one). C, comparison of the results obtained using LASSO and Elastic-NET regressions. On botton, the model generated using all subjects from dataset 2A and the results obtained using this model on dataset 2B.

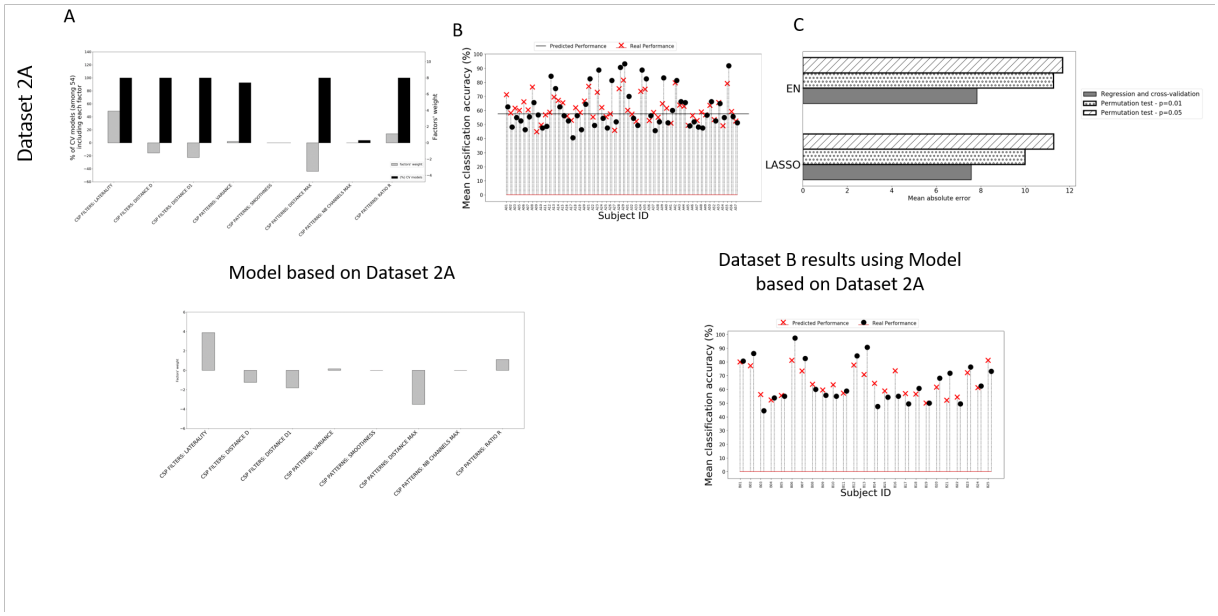


Figure 13.1: Results of the different models generated using an Elastic Net regression and all CSP/LDA-related characteristics (see Chapter 9, Model 1). A, the percentage of Cross-Validation models including each factor (black) and their weight (grey). B, in black (circle), the real performance of each subject and in red (cross), the predicted performance of each subject generated using the model generated from the training dataset (All subjects except the target one). C, comparison of the results obtained using LASSO and Elastic-NET regressions. On botton, the model generated using all subjects from dataset 2A and the results obtained using this model on dataset 2B.

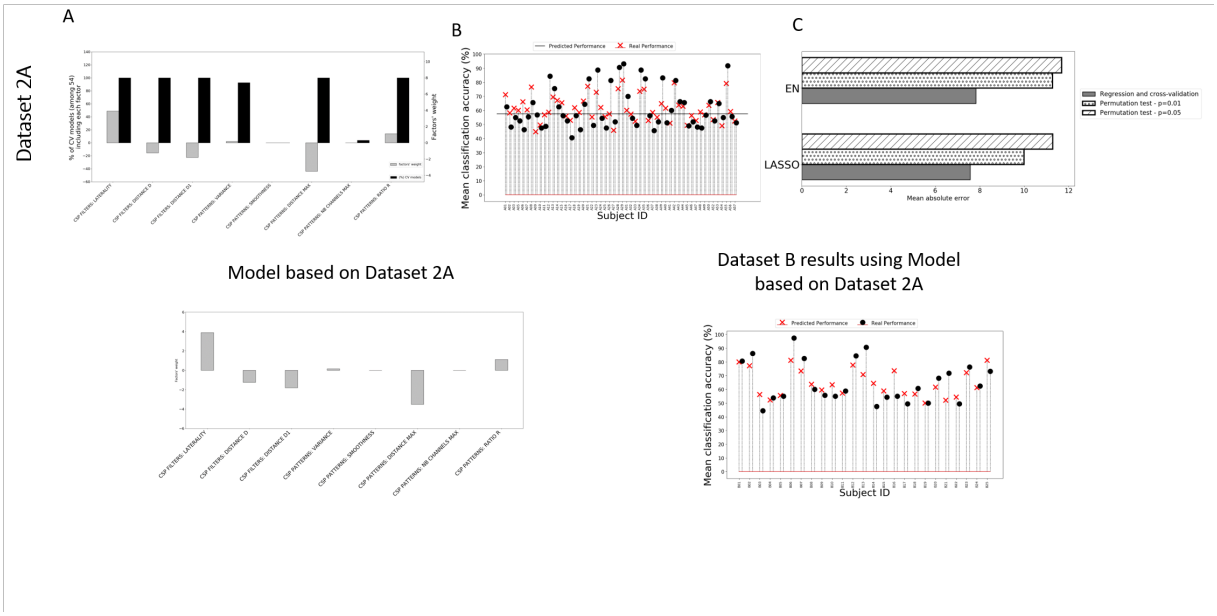


Figure 13.2: Results of the different models generated using an Elastic Net regression and CSP-related characteristics weighted by the LDA patterns normalized weights (see Chapter 9, Model 2). A, the percentage of Cross-Validation models including each factor (black) and their weight (grey). B, in black (circle), the real performance of each subject and in red (cross), the predicted performance of each subject generated using the model generated from the training dataset (All subjects except the target one). C, comparison of the results obtained using LASSO and Elastic-NET regressions. On botton, the model generated using all subjects from dataset 2A and the results obtained using this model on dataset 2B.

Bibliography

[1] M. Clerc, L. Bougrain, and F. Lotte. *Brain-Computer Interfaces 1: Foundations and Methods*. ISTE-Wiley, 2016.

[2] M. Clerc, L. Bougrain, and F. Lotte. *Interfaces Cerveau-Ordinateur 2 : Technologie et Applications*. ISTE-Wiley, 2016.

[3] Benjamin Blankertz, Claudia Sannelli, Sebastian Halder, Eva M Hammer, Andrea Kübler, Klaus-Robert Müller, Gabriel Curio, and Thorsten Dickhaus. Neurophysiological predictor of smr-based bci performance. *Neuroimage*, 51(4):1303–1309, 2010.

[4] Brendan Z Allison and Christa Neuper. Could anyone use a bci? In *Brain-computer interfaces*, pages 35–54. Springer, 2010.

[5] Fabien Lotte, Laurent Bougrain, Andrzej Cichocki, Maureen Clerc, Marco Congedo, Alain Rakotomamonjy, and Florian Yger. A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update. *Journal of neural engineering*, 15(3):031005, 2018.

[6] Fabien Lotte, Florian Larrue, and Christian Mühl. Flaws in current human training protocols for spontaneous brain-computer interfaces: lessons learned from instructional design. *Frontiers in human neuroscience*, 7:568, 2013.

[7] Fabien Lotte and Camille Jeunet. Towards improved bci based on human learning principles. pages 1–4, 2015.

[8] Camille Jeunet, Emilie Jahanpour, and Fabien Lotte. Why standard brain-computer interface (bci) training protocols should be changed: an experimental study. *Journal of neural engineering*, 13(3):036024, 2016.

[9] Jelena Mladenović, Jérémie Mattout, and Fabien Lotte. A generic framework for adaptive eeg-based bci training and operation. *arXiv preprint arXiv:1707.07935*, 2017.

[10] Christa Neuper, Reinhold Scherer, Selina Wriessnegger, and Gert Pfurtscheller. Motor imagery and action observation: modulation of sensorimotor brain rhythms during mental control of a brain-computer interface. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 120(2):239–247, February 2009.

[11] Christa Neuper and Gert Pfurtscheller. Neurofeedback training for bci control. *Brain-computer interfaces*, pages 65–78, 2009.

[12] Aline Roc, Léa Pillette, Jelena Mladenovic, Camille Benaroch, Bernard N'Kaoua, Camille Jeunet, and Fabien Lotte. A review of user training methods in brain computer interfaces based on mental tasks. *Journal of Neural Engineering*, 2020.

[13] C. Jeunet, B. N'Kaoua, and F. Lotte. Advances in user-training for mental-imagery-based bci control: Psychological and cognitive factors and their neural correlates. *Prog. Brain Res.*, 228:3–35, 2016.

[14] Benjamin Blankertz, Claudia Sannelli, Sebastian Halder, Eva M Hammer, Andrea Kübler, Klaus-Robert Müller, Gabriel Curio, and Thorsten Dickhaus. Neurophysiological predictor of SMR-based BCI performance. *Neuroimage*, 51(4):1303–1309, 2010.

[15] Minkyu Ahn and Sung Chan Jun. Performance variation in motor imagery brain–computer interface: a brief review. *Journal of neuroscience methods*, 243:103–110, 2015.

[16] Jonathan R Wolpaw, Niels Birbaumer, Dennis J McFarland, Gert Pfurtscheller, and Theresa M Vaughan. Brain–computer interfaces for communication and control. *Clinical neurophysiology*, 113(6):767–791, 2002.

[17] Herbert Ramoser, Johannes Muller-Gerking, and Gert Pfurtscheller. Optimal spatial filtering of single trial eeg during imagined hand movement. *IEEE transactions on rehabilitation engineering*, 8(4):441–446, 2000.

[18] Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, and Klaus-Robert Muller. Optimizing spatial filters for robust eeg single-trial analysis. *IEEE Signal processing magazine*, 25(1):41–56, 2007.

[19] Benjamin Blankertz, Guido Dornhege, Matthias Krauledat, Klaus-Robert Müller, and Gabriel Curio. The non-invasive berlin brain–computer interface: fast acquisition of effective performance in untrained subjects. *NeuroImage*, 37(2):539–550, 2007.

[20] Hans Berger. Über das elektroenkephalogramm des menschen. *Archiv für psychiatrie und nervenkrankheiten*, 87(1):527–570, 1929.

[21] Lindsay F Haas. Hans berger (1873–1941), richard caton (1842–1926), and electroencephalography. *Journal of Neurology, Neurosurgery & Psychiatry*, 74(1):9–9, 2003.

[22] Gregory Razran. The observable and the inferable conscious in current soviet psychophysiology: Interoceptive conditioning, semantic conditioning, and the orienting reflex. *Psychological review*, 68(2):81, 1961.

[23] Joseph Kamiya. Conscious control of brain waves. *Psychology Today*, 1, 1968.

[24] Volker Straebel and Wilm Thoben. Alvin lucier's music for solo performer: Experimental music beyond sonification. *Organised Sound*, 19:17 – 29, 2014.

[25] Jacques J Vidal. Toward direct brain-computer communication. *Annual review of Biophysics and Bioengineering*, 2(1):157–180, 1973.

[26] Thorsten O Zander and Christian Kothe. Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. *Journal of neural engineering*, 8(2):025005, 2011.

[27] Niels Birbaumer, Thomas Elbert, Anthony G Canavan, and Brigitte Rockstroh. Slow potentials of the cerebral cortex and behavior. *Physiological reviews*, 70(1):1–41, 1990.

[28] Gert Pfurtscheller, Christa Neuper, Christoph Guger, WAHW Harkam, Herbert Ramoser, Alois Schlogl, BAOB Obermaier, and MAPM Pregenzer. Current trends in graz brain-computer interface (bci) research. *IEEE transactions on rehabilitation engineering*, 8(2):216–219, 2000.

[29] Gert Pfurtscheller, Ch Neuper, Doris Flotzinger, and Martin Pregenzer. Eeg-based discrimination between imagination of right and left hand movement. *Electroencephalography and clinical Neurophysiology*, 103(6):642–651, 1997.

[30] Gert Pfurtscheller and Christa Neuper. Motor imagery activates primary sensorimotor area in humans. *Neuroscience letters*, 239(2-3):65–68, 1997.

[31] Jonathan R Wolpaw and Dennis J McFarland. Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proceedings of the national academy of sciences*, 101(51):17849–17854, 2004.

[32] Andrea Kübler, Femke Nijboer, Jürgen Mellinger, Theresa M Vaughan, Hannelore Pawelzik, Gerwin Schalk, Dennis J McFarland, Niels Birbaumer, and Jonathan R Wolpaw. Patients with als can use sensorimotor rhythms to operate a brain-computer interface. *Neurology*, 64(10):1775–1777, 2005.

[33] José del R Millán, Rüdiger Rupp, Gernot Mueller-Putz, Roderick Murray-Smith, Claudio Giugliemma, Michael Tangermann, Carmen Vidaurre, Febo Cincotti, Andrea Kubler, Robert Leeb, et al. Combining brain–computer interfaces and assistive technologies: state-of-the-art and challenges. *Frontiers in neuroscience*, 4:161, 2010.

[34] Anatole Lécuyer, Fabien Lotte, Richard B Reilly, Robert Leeb, Michitaka Hirose, and Mel Slater. Brain-computer interfaces, virtual reality, and videogames. *Computer*, 41(10):66–72, 2008.

[35] Dennis J McFarland and Jonathan R Wolpaw. Brain-computer interfaces for communication and control. *Communications of the ACM*, 54(5):60–66, 2011.

[36] Lawrence Ashley Farwell and Emanuel Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical Neurophysiology*, 70(6):510–523, 1988.

[37] Kevin M Spencer and John Polich. Poststimulus eeg spectral analysis and p300: attention, task, and probability. *Psychophysiology*, 36(2):220–232, 1999.

[38] Gianluca Borghini, Pietro Aricò, Gianluca Di Flumeri, Giulia Cartocci, Alfredo Colosimo, Stefano Bonelli, Alessia Golfetti, Jean Paul Imbert, Géraud Granger, Railane Benhacene, et al. Eeg-based cognitive control behaviour assessment: an ecological study with professional air traffic controllers. *Scientific reports*, 7(1):1–16, 2017.

[39] Jonathan R Wolpaw. Brain–computer interfaces. In *Handbook of Clinical Neurology*, volume 110, pages 67–74. Elsevier, 2013.

[40] Keinosuke Fukunaga. *Introduction to statistical pattern recognition*. Elsevier, 2013.

[41] Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–430, 2000.

[42] William D Penny, Stephen J Roberts, Eleanor A Curran, and Maria J Stokes. Eeg-based communication: a pattern recognition approach. *IEEE transactions on Rehabilitation Engineering*, 8(2):214–215, 2000.

[43] Christopher JC Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167, 1998.

[44] Steven G Mason and Gary E Birch. A general framework for brain-computer interface design. *IEEE transactions on neural systems and rehabilitation engineering*, 11(1):70–85, 2003.

[45] Jonathan R Wolpaw, Gerald E Loeb, Brendan Z Allison, Emanuel Donchin, Omar Feix do Nascimento, William J Heetderks, Femke Nijboer, William G Shain, and James N Turner. Bci meeting 2005-workshop on signals and recording methods. *IEEE Transactions on neural systems and rehabilitation engineering*, 14(2):138–141, 2006.

[46] Ali Bashashati, Mehrdad Fatourechi, Rabab K Ward, and Gary E Birch. A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals. *Journal of Neural engineering*, 4(2):R32, 2007.

[47] Mohd Zaizu Ilyas, Puteh Saad, and Muhammad Imran Ahmad. A survey of analysis and classification of eeg signals for brain-computer interfaces. pages 1–6, 2015.

[48] Jonathan R Wolpaw, Dennis J McFarland, and Theresa M Vaughan. Brain-computer interface research at the wadsworth center. *IEEE Transactions on Rehabilitation Engineering*, 8(2):222–226, 2000.

[49] Gert Pfurtscheller and Christa Neuper. Motor imagery and direct brain-computer communication. *Proceedings of the IEEE*, 89(7):1123–1134, 2001.

[50] Benjamin Blankertz, Guido Dornhege, Matthias Krauledat, K-R Muller, Volker Kunzmann, Florian Losch, and Gabriel Curio. The berlin brain-computer interface: Eeg-based communication without subject training. *IEEE transactions on neural systems and rehabilitation engineering*, 14(2):147–152, 2006.

[51] Jose del R Millan, Josep Mouriño, Marco Franzé, Febo Cincotti, Markus Varsta, Jukka Heikkonen, and Fabio Babiloni. A local neural classifier for the recognition of eeg patterns associated to mental tasks. *IEEE transactions on neural networks*, 13(3):678–686, 2002.

[52] C Jeunet, B N'Kaoua, S Subramanian, M Hachet, and F Lotte. Predicting mental imagery-based bci performance from personality, cognitive profile and neurophysiological patterns. *PloS one*, 10(12):e0143962, 2015.

[53] A. Kübler, B. Kotchoubey, J. Kaiser, J. Wolpaw, and N. Birbaumer. Brain-computer communication: Unlocking the locked in. *Psychological Bulletin*, 127(3):358–375, 2001.

[54] María A Cervera, Surjo R Soekadar, Junichi Ushiba, José del R Millán, Meigen Liu, Niels Birbaumer, and Gangadhar Garipelli. Brain-computer interfaces for post-stroke motor rehabilitation: a meta-analysis. *Annals of clinical and translational neurology*, 5(5):651–663, 2018.

[55] Maureen Clerc, Laurent Bougrain, and Fabien Lotte. Brain-computer interfaces: Foundations and methods. 2016.

[56] Thomas C Ferree, MT Clay, and Don M Tucker. The spatial resolution of scalp eeg. *Neurocomputing*, 38:1209–1216, 2001.

[57] Margaret C. Thompson. Critiquing the concept of BCI illiteracy. *Science and Engineering Ethics*, 25(4):1217–1233, 2018.

[58] Bernhard Graimann, Brendan Z Allison, and Gert Pfurtscheller. *Brain-computer interfaces: Revolutionizing human-computer interaction*. Springer Science & Business Media, 2010.

[59] William F Chaplin, Oliver P John, and Lewis R Goldberg. Conceptions of states and traits: dimensional attributes with ideals as prototypes. *Journal of personality and social psychology*, 54(4):541, 1988.

[60] Wenke Burde and Benjamin Blankertz. Is the locus of control of reinforcement a predictor of brain-computer interface performance? 2006.

[61] Eva Maria Hammer, Sebastian Halder, Benjamin Blankertz, Claudia Sannelli, Thorsten Dickhaus, Sonja Kleih, Klaus-Robert Müller, and Andrea Kübler. Psychological predictors of smr-bci performance. *Biological psychology*, 89(1):80–86, 2012.

[62] Irene Daum, Brigitte Rockstroh, Niels Birbaumer, Thomas Elbert, Anthony Canavan, and Werner Lutzenberger. Behavioural treatment of slow cortical potentials in intractable epilepsy: neuropsychological predictors of outcome. *Journal of Neurology, Neurosurgery & Psychiatry*, 56(1):94–97, 1993.

[63] Eva M Hammer, Tobias Kaufmann, Sonja C Kleih, Benjamin Blankertz, and Andrea Kübler. Visuo-motor coordination ability predicts performance with brain-computer interfaces controlled by modulation of sensorimotor rhythms (smr). *Frontiers in human neuroscience*, 8:574, 2014.

[64] Camille Jeunet. Understanding & improving mental-imagery based brain-computer interface (mi-bci) user-training : towards a new generation of reliable, efficient & accessible brain-computer interfaces.

[65] K Pacheco, K Acuña, E Carranza, D Achanccaray, and J Andreu-Perez. Performance predictors of motor imagery brain-computer interface based on spatial abilities for upper limb rehabilitation. In *Proc. IEEE EMBC*, pages 1014–1017. IEEE, 2017.

[66] Aleksandra Vuckovic and Bethel A Osuagwu. Using a motor imagery questionnaire to estimate the performance of a brain–computer interface based on object oriented motor imagery. *Clinical Neurophysiology*, 124(8):1586–1595, 2013.

[67] Adriane B Randolph. Not all created equal: individual-technology fit of brain-computer interfaces. In *2012 45th Hawaii International Conference on System Sciences*, pages 572–578. IEEE, 2012.

[68] Dariusz Zapała, Emilia Zabielska-Mendyk, Paweł Augustynowicz, Andrzej Cudo, Marta Jaśkiewicz, Marta Szewczyk, Natalia Kopiś, and Piotr Francuz. The effects of handedness on sensorimotor rhythm desynchronization and motor-imagery bci control. *Scientific reports*, 10(1):1–11, 2020.

[69] Moritz Grosse-Wentrup, Bernhard Schölkopf, and Jeremy Hill. Causal influence of gamma oscillations on the sensorimotor rhythm. *NeuroImage*, 56(2):837–842, 2011.

[70] Nicola Neumann and Niels Birbaumer. Predictors of successful self control during brain-computer communication. *Journal of Neurology, Neurosurgery & Psychiatry*, 74(8):1117–1121, 2003.

[71] Femke Nijboer, EW Sellers, Jürgen Mellinger, Mary Ann Jordan, Tamara Matuz, Adrian Furdea, Sebastian Halder, Ursula Mochty, DJ Krusienski, TM Vaughan, et al. A p300-based brain–computer interface for people with amyotrophic lateral sclerosis. *Clinical neurophysiology*, 119(8):1909–1916, 2008.

[72] Matthias Witte, Silvia Erika Kober, Manuel Ninaus, Christa Neuper, and Guilherme Wood. Control beliefs can predict the ability to up-regulate sensorimotor rhythm during neurofeedback training. *Frontiers in human neuroscience*, 7:478, 2013.

[73] AB Randolph, MM Jackson, and S Karmakar. Individual characteristics and their effect on predicting mu rhythm modulation. *Int J HUM-COMPUT INT*, 27(1):24–37, 2010.

[74] Atieh Bamdadian, Cuntai Guan, Kai Keng Ang, and Jianxin Xu. The predictive role of pre-cue eeg rhythms on mi-based bci classification performance. *Journal of neuroscience methods*, 235:138–144, 2014.

[75] Moritz Grosse-Wentrup and Bernhard Schölkopf. High gamma-power predicts performance in sensorimotor-rhythm brain–computer interfaces. *Journal of neural engineering*, 9(4):046001, 2012.

[76] Wolfgang Klimesch. Eeg alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews*, 29(2-3):169–195, 1999.

[77] Alexander Ya Kaplan, Andrew A Fingelkurts, Alexander A Fingelkurts, Sergei V Borisov, and Boris S Darkhovsky. Nonstationary nature of the brain activity as revealed by eeg/meg: methodological, practical and conceptual challenges. *Signal processing*, 85(11):2190–2212, 2005.

[78] Moritz Grosse-Wentrup and Bernhard Schölkopf. A review of performance variations in smr-based brain- computer interfaces (bcis). *Brain-Computer Interface Research*, pages 39–51, 2013.

[79] Ricardo Chavarriaga, Melanie Fried-Oken, Sonja Kleih, Fabien Lotte, and Reinhold Scherer. Heading for new shores! overcoming pitfalls in BCI design. *Brain-Computer Interfaces*, 4(1-2):60–73, 2016.

[80] C Guger, G Edlinger, W Harkam, I Niedermayer, and G Pfurtscheller. How many people are able to operate an eeg-based brain-computer interface (bci)? *IEEE transactions on neural systems and rehabilitation engineering*, 11(2):145–147, 2003.

[81] Brendan Allison, Thorsten Luth, Diana Valbuena, Amir Teymourian, Ivan Volosyak, and Axel Graser. Bci demographics: How many (and what kinds of) people can use an ssvep bci? *IEEE transactions on neural systems and rehabilitation engineering*, 18(2):107–116, 2010.

[82] Fabien Lotte, Marco Congedo, Anatole Lécuyer, Fabrice Lamarche, and Bruno Arnaldi. A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of neural engineering*, 4(2):R1, 2007.

[83] Ian Daly, Martin Billinger, José Laparra-Hernández, Fabio Aloise, Mariano Lloria García, Josef Faller, Reinhold Scherer, and Gernot Müller-Putz. On the control of brain-computer interfaces by users with cerebral palsy. *Clinical Neurophysiology*, 124(9):1787–1797, 2013.

[84] Laura Kauhanen, Pasi Jylänki, Janne Lehtonen, Pekka Rantanen, Hannu Alaranta, and Mikko Sams. EEG-based brain-computer interface for tetraplegics. *Computational intelligence and neuroscience*, 2007, 2007.

[85] Jan Conradi, Benjamin Blankertz, Michael Tangermann, Volker Kunzmann, and Gabriel Curio. Brain-computer interfacing in tetraplegic patients with high spinal cord injury. *Int. J. Bioelectromagn*, 11(2):65–68, 2009.

[86] Stephanie Brandl, Johannes Höhne, Klaus-Robert Müller, and Wojciech Samek. Bringing BCI into everyday life: Motor imagery in a pseudo realistic environment. In *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 224–227. IEEE, 2015.

[87] Karina Statthaler, Andreas Schwarz, David Steyrl, Reinmar Kobler, Maria Katharina Höller, Julia Brandstetter, Lea Hehenberger, Marvin Bigga, and Gernot Müller-Putz. Cybathlon experiences of the Graz BCI racing team Mirage91 in the brain-computer interface discipline. *Journal of neuroengineering and rehabilitation*, 14(1):1–16, 2017.

[88] Serafeim Perdikis, Luca Tonin, Sareh Saeedi, Christoph Schneider, and José del R Millán. The Cybathlon BCI race: Successful longitudinal mutual learning with two tetraplegic users. *PLoS biology*, 16(5):e2003787, 2018.

[89] Serafeim Perdikis and Jose del R Millan. Brain-machine interfaces: A tale of two learners. *IEEE Systems, Man, and Cybernetics Magazine*, 6(3):12–19, 2020.

[90] Domen Novak, Roland Sigrist, Nicolas J Gerig, Dario Wyss, René Bauer, Ulrich Götz, and Robert Riener. Benchmarking brain-computer interfaces outside the laboratory: The cybathlon 2016. *Frontiers in neuroscience*, 11:756, 2018.

[91] Ricardo Chavarriaga, Melanie Fried-Oken, Sonja Kleih, Fabien Lotte, and Reinhold Scherer. Heading for new shores! overcoming pitfalls in BCI design. *Brain-Computer Interfaces*, 4(1-2):60–73, 2016.

[92] Dennis J McFarland, William A Sarnacki, and Jonathan R Wolpaw. Electroencephalographic (EEG) control of three-dimensional movement. *Journal of neural engineering*, 7(3):036007, 2010.

[93] Yan Li, Hiroyuki Kambara, Yasuharu Koike, and Masashi Sugiyama. Application of covariate shift adaptation techniques in brain–computer interfaces. *IEEE Transactions on Biomedical Engineering*, 57(6):1318–1324, 2010.

[94] Pradeep Shenoy, Matthias Krauledat, Benjamin Blankertz, Rajesh PN Rao, and Klaus-Robert Müller. Towards adaptive classification for BCI. *Journal of neural engineering*, 3(1):R13, 2006.

[95] Florian Yger, Maxime Berar, and Fabien Lotte. Riemannian approaches in brain-computer interfaces: a review. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(10):1753–1762, 2017.

[96] Marco Congedo, Alexandre Barachant, and Rajendra Bhatia. Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review. *Brain-Computer Interfaces*, 4(3):155–174, 2017.

[97] Maria Yamamoto, Khadijeh Sadatnejad, Toshihisa Tanaka, Md Rabiul Islam, Yuichi Tanaka, and Fabien Lotte. Detecting EEG outliers for BCI on the riemannian manifold using spectral clustering. 2020.

[98] Fabien Lotte, Laurent Bougrain, Andrzej Cichocki, Maureen Clerc, Marco Congedo, Alain Rakotomamonjy, and Florian Yger. A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *Journal of neural engineering*, 15(3):031005, 2018.

[99] Satyam Kumar, Florian Yger, and Fabien Lotte. Towards adaptive classification using riemannian geometry approaches in brain-computer interfaces. In *2019 7th International Winter Conference on Brain-Computer Interface (BCI)*, pages 1–6. IEEE, 2019.

[100] Elisabeth VC Friedrich, Christa Neuper, and Reinhold Scherer. Whatever works: a systematic user-centered training protocol to optimize brain-computer interfacing individually. *PloS one*, 8(9):e76214, 2013.

[101] M David Merrill. First principles of instruction: A synthesis. *Trends and issues in instructional design and technology*, 2:62–71, 2007.

[102] Yann Renard, Fabien Lotte, Guillaume Gibert, Marco Congedo, Emmanuel Maby, Vincent Delannoy, Olivier Bertrand, and Anatole Lécuyer. Openvibe: An open-source software platform to design, test, and use brain–computer interfaces in real and virtual environments. *Presence*, 19(1):35–53, 2010.

[103] Gert Pfurtscheller and Christa Neuper. Motor imagery and direct brain-computer communication. *Proceedings of the IEEE*, 89(7):1123–1134, 2001.

[104] Aline Roc, Léa Pillette, Bernard N'Kaoua, and Fabien Lotte. Would motor-imagery based BCI user training benefit from more women experimenters? In *Int. Graz BCI Conf.*, 2019.

[105] Léa Pillette, Aline Roc, Bernard N'Kaoua, and Fabien Lotte. Experimenters' Influence on MI-BCI User Training. *International Journal of Human-Computer Studies, in press*, 2021.

[106] Julie Bismuth, François Vialatte, and Jean-Pascal Lefaucheur. Relieving peripheral neuropathic pain by increasing the power-ratio of low-β over high-β activities in the central cortical region with eeg-based neurofeedback: study protocol for a controlled pilot trial (smrpain study). *Neurophysiologie Clinique*, 50(1):5–20, 2020.

[107] Viswanath Venkatesh and Hillol Bala. Technology acceptance model 3 and a research agenda on interventions. *Decision sciences*, 39(2):273–315, 2008.

[108] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. John Wiley & Sons, 2012.

[109] Mahnaz Arvaneh, Cuntai Guan, Kai Keng Ang, and Chai Quek. Robust EEG channel selection across sessions in brain-computer interface involving stroke patients. In *The 2012 International Joint Conference on Neural Networks (IJCNN)*, pages 1–6. IEEE, 2012.

[110] Boris Reuderink and Mannes Poel. Robustness of the common spatial patterns algorithm in the BCI-pipeline. *University of Twente, Tech. Rep*, 2008.

[111] Fabien Lotte. A tutorial on EEG signal-processing techniques for mental-state recognition in brain–computer interfaces. In *Guide to Brain-Computer Music Interfacing*, pages 133–161. Springer, 2014.

[112] Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, 88(2):365–411, 2004.

[113] Serge Lang. Differential and riemannian manifolds. 160, 2012.

[114] Maher Moakher. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 26(3):735–747, 2005.

[115] Alexandre Barachant, Stéphane Bonnet, Marco Congedo, and Christian Jutten. Riemannian geometry applied to BCI classification. In *International Conference on Latent Variable Analysis and Signal Separation*, pages 629–636. Springer, 2010.

[116] Paolo Zanini, Marco Congedo, Christian Jutten, Salem Said, and Yannick Berthoumieu. Transfer learning: A riemannian geometry framework with applications to brain–computer interfaces. *IEEE Transactions on Biomedical Engineering*, 65(5):1107–1116, 2017.

[117] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[118] Reinhold Scherer, Josef Faller, Paul Sajda, and Carmen Vidaurre. Eeg-based endogenous online co-adaptive brain-computer interfaces: Strategy for success? pages 299–304, 2018.

[119] Fabien Lotte and Camille Jeunet. Defining and quantifying users' mental imagery-based bci skills: a first step. *Journal of neural engineering*, 15, 2018.

[120] Serafeim Perdikis, Robert Leeb, and J d R Millán. Context-aware adaptive spelling in motor imagery BCI. *Journal of neural engineering*, 13(3):036018, 2016.

[121] F Chochon, L Cohen, PF van de Moortele, and Stanislas Dehaene. Differential contributions of the left and right inferior parietal lobules to number processing. *Journal of cognitive neuroscience*, 11(6):617–630, 1999.

[122] G Müller-Putz, R Scherer, C Brunner, R Leeb, and G Pfurtscheller. Better than random: a closer look on bci results. *Int J Bioelectromagn*, 10:52–55, 2008.

[123] John Sweller, Jeroen JG van Merriënboer, and Fred Paas. Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, pages 1–32, 2019.

[124] Hwan-Hee Choi, Jeroen JG Van Merriënboer, and Fred Paas. Effects of the physical environment on cognitive load and learning: towards a new model of cognitive load. *Educational Psychology Review*, 26(2):225–244, 2014.

[125] R Tibshirani. Regression shrinkage and selection via the lasso. *J R Stat Soc Series B Stat Methodol*, pages 267–288, 1996.

[126] RB Cattell and HE P. Cattell. Personality structure and the new fifth edition of the 16pf. *Educ Psychol Meas*, 55(6):926–937, 1995.

[127] JM ALBARET and E AUBERT. Etalonnage 15-19 ans du test de rotation mentale de vandenberg. *Evolutions psychomotrices*, (34):206–209, 1996.

[128] Steven G Vandenberg and Allan R Kuse. Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and motor skills*, 47(2):599–604, 1978.

[129] Roger N Shepard and Jacqueline Metzler. Mental rotation of three-dimensional objects. *Science*, 171(3972):701–703, 1971.

[130] Léa Pillette, Aline Roc, Bernard N'kaoua, and Fabien Lotte. Experimenters' influence on mental-imagery based brain-computer interface user training. *International Journal of Human-Computer Studies*, 149:102603, 2021.

[131] Maria Yamamoto, Camille Benaroch, Aline Roc, Thibaut Monseigne, and Fabien Lotte. Should frequency band selection algorithms include neurophysiological constraints? 2021.

[132] S Teillet, F Lotte, B N'Kaoua, and C Jeunet. Towards a spatial ability training to improve mental imagery based brain-computer interface (mi-bci) performance: A pilot study. In *Proc. IEEE SMC*, pages 003664–003669, 2016.

[133] Léa Pillette, Camille Jeunet, Boris Mansencal, Roger N'kambou, Bernard N'Kaoua, and Fabien Lotte. A physical learning companion for mental-imagery BCI user training. *International Journal of Human-Computer Studies*, 136:102380, 2020.

[134] Elisabeth VC Friedrich, Reinhold Scherer, and Christa Neuper. Long-term evaluation of a 4-class imagery-based brain–computer interface. *Clinical Neurophysiology*, 124(5):916–927, 2013.

[135] Silvia Erika Kober, Matthias Witte, Manuel Ninaus, Christa Neuper, and Guilherme Wood. Learning to modulate one's own brain activity: the effect of spontaneous mental strategies. *Frontiers in human neuroscience*, 7:695, 2013.

[136] Christa Neuper, Reinhold Scherer, Miriam Reiner, and Gert Pfurtscheller. Imagery of motor actions: Differential effects of kinesthetic and visual–motor mode of imagery in single-trial eeg. *Cognitive brain research*, 25(3):668–677, 2005.

[137] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans Neural Syst Rehabil Eng*, 8(4):441–446, 2000.

[138] Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, and Klaus-Robert Muller. Optimizing spatial filters for robust eeg single-trial analysis. *IEEE Signal processing magazine*, 25(1):41–56, 2007.

[139] F Lotte. Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces. *Proc. IEEE*, (103):871–890, 2015.

[140] V Fonti and E Belitser. Feature selection using lasso. *VU Amsterdam Research Paper in Business Analytics*, 2017.

[141] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2):301–320, 2005.

[142] Nikki Leeuwis and Maryam Alimardani. High aptitude motor-imagery bci users have better visuospatial memory. pages 1518–1523, 2020.

[143] Nikki Leeuwis, Alissa Paas, and Maryam Alimardani. Vividness of visual imagery and personality impact motor-imagery brain computer interfaces. *Frontiers in Human Neuroscience*, 15, 2021.

[144] Loic Botrel and Andrea Kübler. Reliable predictors of smr bci performance—do they exist? pages 1–3, 2018.

[145] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, 1995.

[146] DA Kolb. *Learning style inventory: Version 3*. Hay/McBer Training Resources Group, 1999.

[147] Fabien Lotte, Camille Jeunet, Ricardo Chavarriaga, Laurent Bougrain, Dave E Thompson, Reinhold Scherer, Md Rakibul Mowla, Andrea Kübler, Moritz Grosse-Wentrup, Karen Dijkstra, and Natalie Dayan. Turning negative into positives! exploiting 'negative'results in brain–machine interface (BMI) research. *Brain-Computer Interfaces*, 6(4):178–189, 2019.

[148] H Xu, C Caramanis, and S Mannor. Robust regression and lasso. pages 1801–1808, 2009.

[149] Loann David Denis Desboulets. A review on variable selection in regression analysis. *Econometrics*, 6(4):45, 2018.

[150] J Craig Henry. Electroencephalography: basic principles, clinical applications, and related fields. *Neurology*, 67(11):2092–2092, 2006.

[151] Richard W Homan, John Herman, and Phillip Purdy. Cerebral location of international 10–20 system electrode placement. *Electroencephalography and clinical neurophysiology*, 66(4):376–382, 1987.

[152] Jayant N Acharya, Abeer J Hani, Janna Cheek, Parthasarathy Thirumala, and Tammy N Tsuchida. American clinical neurophysiology society guideline 2: guidelines for standard electrode position nomenclature. *The Neurodiagnostic Journal*, 56(4):245–252, 2016.

[153] Michal Teplan et al. Fundamentals of eeg measurement. *Measurement science review*, 2(2):1–11, 2002.

[154] Richard B Berry, Rita Brooks, Charlene E Gamaldo, Susan M Harding, C Marcus, Bradley V Vaughn, et al. The aasm manual for the scoring of sleep and associated events. *Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine*, 176:2012, 2012.

[155] Stefanie Liebe, Gregor M Hoerzer, Nikos K Logothetis, and Gregor Rainer. Theta coupling between v4 and prefrontal cortex predicts visual short-term memory performance. *Nature neuroscience*, 15(3):456–462, 2012.

[156] Ueli Rutishauser, Ian B Ross, Adam N Mamelak, and Erin M Schuman. Human memory strength is predicted by theta-frequency phase-locking of single neurons. *Nature*, 464(7290):903–907, 2010.

[157] Constantine Pavlides, Yoram J Greenstein, Mark Grudman, and Jonathan Winson. Long-term potentiation in the dentate gyrus is induced preferentially on the positive phase of θ-rhythm. *Brain research*, 439(1-2):383–387, 1988.

[158] G Orr, G Rao, FP Houston, BL McNaughton, and Carol A Barnes. Hippocampal synaptic plasticity is modulated by theta rhythm in the fascia dentata of adult and aged freely behaving rats. *Hippocampus*, 11(6):647–654, 2001.

[159] Avishek Adhikari, Mihir A Topiwala, and Joshua A Gordon. Synchronized activity between the ventral hippocampus and the medial prefrontal cortex during anxiety. *Neuron*, 65(2):257–269, 2010.

[160] Dieter Vaitl, Niels Birbaumer, John Gruzelier, Graham A Jamieson, Boris Kotchoubey, Andrea Kübler, Dietrich Lehmann, Wolfgang HR Miltner, Ulrich Ott, Peter Pütz, et al. Psychobiology of altered states of consciousness. *Psychological bulletin*, 131(1):98, 2005.

[161] Gregg D Jacobs and Richard Friedman. Eeg spectral analysis of relaxation techniques. *Applied psychophysiology and biofeedback*, 29(4):245–254, 2004.

[162] Alfred L Loomis, E Newton Harvey, and Garret A Hobart III. Distribution of disturbance-patterns in the human electroencephalogram, with special reference to sleep. *Journal of Neurophysiology*, 1(5):413–430, 1938.

[163] Mario Rosanova, Adenauer Casali, Valentina Bellina, Federico Resta, Maurizio Mariotti, and Marcello Massimini. Natural frequencies of human corticothalamic circuits. *Journal of Neuroscience*, 29(24):7679–7685, 2009.

[164] Wolf Singer and Charles M Gray. Visual feature integration and the temporal correlation hypothesis. *Annual review of neuroscience*, 18(1):555–586, 1995.

[165] Andreas K Engel, Peter König, Andreas K Kreiter, Thomas B Schillen, and Wolf Singer. Temporal coding in the visual cortex: new vistas on integration in the nervous system. *Trends in neurosciences*, 15(6):218–226, 1992.

[166] Ole Jensen, Jochen Kaiser, and Jean-Philippe Lachaux. Human gamma-frequency oscillations associated with attention and memory. *Trends in neurosciences*, 30(7):317–324, 2007.

[167] Floris P De Lange, Ole Jensen, Markus Bauer, and Ivan Toni. Interactions between posterior gamma and frontal alpha/beta oscillations during imagined actions. *Frontiers in human neuroscience*, 2:7, 2008.

[168] Giacomo Rizzolatti, Luigi Cattaneo, Maddalena Fabbri-Destro, and Stefano Rozzi. Cortical mechanisms underlying the organization of goal-directed actions and mirror neuron-based action understanding. *Physiological reviews*, 94(2):655–706, 2014.

[169] Giovanni Buccino and Lucia Riggio. The role of the mirror neuron system in motor learning. *Kinesiology*, 38(1.):5–15, 2006.

[170] Riitta Hari and Riitta Salmelin. Human cortical oscillations: a neuromagnetic view through the skull. *Trends in neurosciences*, 20(1):44–49, 1997.

[171] Dennis J McFarland, Laurie A Miner, Theresa M Vaughan, and Jonathan R Wolpaw. Mu and beta rhythm topographies during motor imagery and actual movements. *Brain topography*, 12(3):177–186, 2000.

[172] Gert Pfurtscheller and A Aranibar. Evaluation of event-related desynchronization (erd) preceding and following voluntary self-paced movement. *Electroencephalography and clinical neurophysiology*, 46(2):138–146, 1979.

[173] Gert Pfurtscheller and FH Lopes Da Silva. Event-related eeg/meg synchronization and desynchronization: basic principles. *Clinical neurophysiology*, 110(11):1842–1857, 1999.

[174] Heung-Il Suk, Siamac Fazli, Jan Mehnert, Klaus-Robert Müller, and Seong-Whan Lee. Predicting bci subject performance using probabilistic spatio-temporal filters. *PloS one*, 9(2):e87056, 2014.

[175] Giovanni Buccino, Fausta Lui, Nicola Canessa, Ilaria Patteri, Giovanna Lagravinese, Francesca Benuzzi, Carlo A Porro, and Giacomo Rizzolatti. Neural circuits involved in the recognition of actions performed by nonconspecifics: An fmri study. *Journal of cognitive neuroscience*, 16(1):114–126, 2004.

[176] Beatriz Calvo-Merino, Daniel E Glaser, Julie Grèzes, Richard E Passingham, and Patrick Haggard. Action observation and acquired motor skills: an fmri study with expert dancers. *Cerebral cortex*, 15(8):1243–1249, 2005.

[177] Minkyu Ahn, Hohyun Cho, Sangtae Ahn, and Sung Chan Jun. High theta and low alpha powers may be indicative of bci-illiteracy in motor imagery. *PloS one*, 8(11):e80886, 2013.

[178] Moonyoung Kwon, Hohyun Cho, Kyungho Won, Minkyu Ahn, and Sung Chan Jun. Use of both eyes-open and eyes-closed resting states may yield a more robust predictor of motor imagery bci performance. *Electronics*, 9(4):690, 2020.

[179] Bradley J Edelman, Jianjun Meng, Daniel Suma, C Zurn, E Nagarajan, BS Baxter, Christopher C Cline, and Bin He. Noninvasive neuroimaging enhances continuous neural tracking for robotic device control. *Science robotics*, 4(31), 2019.

[180] Tzyy-Ping Jung, Colin Humphries, Te-Won Lee, Scott Makeig, Martin McKeown, Vicente Iragui, and Terrence J Sejnowski. Extended ica removes artifacts from electroencephalographic recordings. *Advances in neural information processing systems*, 10:894–900, 1997.

[181] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, and Matti S. Hämäläinen. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7(267):1–13, 2013.

[182] Aapo Hyvärinen. Independent component analysis: recent advances. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1984):20110534, 2013.

[183] Mainak Jas, Denis Engemann, Federico Raimondo, Yousra Bekhti, and Alexandre Gramfort. Automated rejection and repair of bad trials in meg/eeg. pages 1–4, 2016.

[184] Israel Gohberg, Seymour Goldberg, and Marinus A Kaashoek. Hilbert-schmidt operators. In *Classes of Linear Operators Vol. I*, pages 138–147. Springer, 1990.

[185] Nima Bigdely-Shamlo, Tim Mullen, Christian Kothe, Kyung-Min Su, and Kay A Robbins. The prep pipeline: standardized preprocessing for large-scale eeg analysis. *Frontiers in neuroinformatics*, 9:16, 2015.

[186] B. Blankertz, C. Sannelli, S. Halder, E. M. Hammer, A. Kübler, K.-R. Müller, G. Curio, , and T. Dickhaus. Neurophysiological predictor of SMR-based BCI performance. *NeuroImage*, 51(4), 2010.

[187] Dennis J McFarland, Lynn M McCane, Stephen V David, and Jonathan R Wolpaw. Spatial filter selection for eeg-based communication. *Electroencephalography and clinical Neurophysiology*, 103(3):386–394, 1997.

[188] Bradley Voytek and Robert T Knight. Dynamic network communication as a unifying neural basis for cognition, development, aging, and disease. *Biological psychiatry*, 77(12):1089–1097, 2015.

[189] Bradley Voytek, Mark A Kramer, John Case, Kyle Q Lepage, Zechari R Tempesta, Robert T Knight, and Adam Gazzaley. Age-related changes in 1/f neural electrophysiological noise. *Journal of Neuroscience*, 35(38):13257–13265, 2015.

[190] Farran Briggs, George R Mangun, and W Martin Usrey. Attention enhances synaptic efficacy and the signal-to-noise ratio in neural circuits. *Nature*, 499(7459):476–480, 2013.

[191] Guang Ouyang, Andrea Hildebrandt, Florian Schmitz, and Christoph S Herrmann. Decomposing alpha and 1/f brain activities reveals their differential associations with cognitive processing speed. *NeuroImage*, 205:116304, 2020.

[192] Claudia Sannelli, Mikio Braun, Michael Tangermann, and Klaus-Robert Müller. Estimating noise and dimensionality in bci data sets: Towards illiteracy comprehension. 2008.

[193] John Gruzelier. A theory of alpha/theta neurofeedback, creative performance enhancement, long distance functional connectivity and psychological integration. *Cognitive processing*, 10(1):101–109, 2009.

[194] Tracy Boynton. Applied research using alpha/theta training for enhancing creativity and well-being. *Journal of Neurotherapy*, 5(1-2):5–18, 2001.

[195] Xiyuan Jiang, Emily Lopez, James Stieger, Carol M Greco, and Bin He. Effects of long-term meditation practices on sensorimotor rhythm based bci learning. *Frontiers in neuroscience*, 14:1443, 2020.

[196] Lee-Fan Tan, Zoltan Dienes, Ashok Jansari, and Sing-Yau Goh. Effect of mindfulness meditation on brain–computer interface performance. *Consciousness and cognition*, 23:12–21, 2014.

[197] John Crosley Shaw. Intention as a component of the alpha-rhythm response to mental activity. *International Journal of Psychophysiology*, 24(1-2):7–23, 1996.

[198] Ronald W Schafer. What is a savitzky-golay filter?[lecture notes]. *IEEE Signal processing magazine*, 28(4):111–117, 2011.

[199] J Ross Quinlan. Learning efficient classification procedures and their application to chess end games. pages 463–482, 1983.

[200] Claudia Sannelli, Thorsten Dickhaus, Sebastian Halder, Eva-Maria Hammer, Klaus-Robert Müller, and Benjamin Blankertz. On optimal channel configurations for smr-based brain–computer interfaces. *Brain topography*, 23(2):186–193, 2010.

[201] Steven W Smith et al. The scientist and engineer's guide to digital signal processing. 1997.

[202] Benjamin Blankertz, K-R Muller, Dean J Krusienski, Gerwin Schalk, Jonathan R Wolpaw, Alois Schlogl, Gert Pfurtscheller, Jd R Millan, Michael Schroder, and Niels Birbaumer. The bci competition iii: Validating alternative approaches to actual bci problems. *IEEE transactions on neural systems and rehabilitation engineering*, 14(2):153–159, 2006.

[203] Marco Congedo, Fabien Lotte, and Anatole Lécuyer. Classification of movement intention by spatially filtered electromagnetic inverse solutions. *Physics in Medicine & Biology*, 51(8):1971, 2006.

[204] Marti A. Hearst, Susan T Dumais, Edgar Osuna, John Platt, and Bernhard Scholkopf. Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4):18–28, 1998.

[205] Ronald A Fisher. The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2):179–188, 1936.

[206] Radford M Neal. *Pattern recognition and machine learning*. Taylor & Francis, 2007.

[207] Stefan Haufe, Frank Meinecke, Kai Görgen, Sven Dähne, John-Dylan Haynes, Benjamin Blankertz, and Felix Bießmann. On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage*, 87:96–110, 2014.

[208] Thomas Naselaris, Kendrick N Kay, Shinji Nishimoto, and Jack L Gallant. Encoding and decoding in fmri. *Neuroimage*, 56(2):400–410, 2011.

[209] Fabien Lotte and Cuntai Guan. Spatially regularized common spatial patterns for eeg classification. pages 3712–3715, 2010.

[210] Zhen James Xiang, Yongxin Taylor Xi, Uri Hasson, and Peter J Ramadge. Boosting with spatial regularization. 22:2107–2115, 2009.

[211] Julia Schumacher, Camille Jeunet, and Fabien Lotte. Towards explanatory feedback for user training in brain-computer interfaces. pages 3169–3174, 2015.

[212] Han-Jeong Hwang, Kiwoon Kwon, and Chang-Hwang Im. Neurofeedback-based motor imagery training for brain–computer interface (bci). *Journal of neuroscience methods*, 179(1):150–156, 2009.

[213] Masahito Mihara, Ichiro Miyai, Noriaki Hattori, Megumi Hatakenaka, Hajime Yagura, Teiji Kawano, Masaki Okibayashi, Nobuyoshi Danjo, Akihiro Ishikawa, Yoshihiro Inoue, et al. Neurofeedback using real-time near-infrared spectroscopy enhances motor imagery related cortical activation. *PloS one*, 7(3):e32234, 2012.

[214] Seung-Schik Yoo, Jong-Hwan Lee, Heather O'Leary, Lawrence P Panych, and Ferenc A Jolesz. Neurofeedback fmri-mediated learning and consolidation of regional brain activation during motor imagery. *International journal of imaging systems and technology*, 18(1):69–78, 2008.

[215] Moritz Grosse-Wentrup et al. What are the causes of performance variation in brain-computer interfacing. *International Journal of Bioelectromagnetism*, 13(3):115–116, 2011.

[216] Lawrence G Weiss, Donald H Saklofske, James A Holdnack, and Aurelio Prifitera. Wisc-v assessment and interpretation: Scientist-practitioner perspectives. 2015.

[217] Fabien Lotte, Camille Jeunet, Ricardo Chavarriaga, Laurent Bougrain, Dave E Thompson, Reinhold Scherer, Md Rakibul Mowla, Andrea Kübler, Moritz Grosse-Wentrup, Karen Dijkstra, et al. Turning negative into positives! exploiting 'negative'results in brain–machine interface (bmi) research. *Brain-Computer Interfaces*, 6(4):178–189, 2019.

[218] John PA Ioannidis. Why science is not necessarily self-correcting. *Perspectives on Psychological Science*, 7(6):645–654, 2012.

[219] Fabien Lotte and Camille Jeunet. Online classification accuracy is a poor metric to study mental imagery-based bci user learning: an experimental demonstration and new metrics. 2017.

[220] Jin Woo Choi, Sejoon Huh, and Sungho Jo. Improving performance in motor imagery bci-based control applications via virtually embodied feedback. *Computers in Biology and Medicine*, 127:104079, 2020.

[221] Michele Barsotti, Daniele Leonardis, Nicola Vanello, Massimo Bergamasco, and Antonio Frisoli. Effects of continuous kinaesthetic feedback based on tendon vibration on motor imagery bci performance. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(1):105–114, 2017.

[222] Álvaro Barbero and Moritz Grosse-Wentrup. Biased feedback in brain-computer interfaces. *Journal of neuroengineering and rehabilitation*, 7(1):34, 2010.

[223] Maryam Alimardani, Shuichi Nishio, and Hiroshi Ishiguro. Effect of biased feedback on motor imagery learning in bci-teleoperation system. *Frontiers in systems neuroscience*, 8:52, 2014.

[224] Camille Jeunet, Camille Bernaroch, F Cabestaing, Ricardo Chavarriaga, Emma Colamarino, Marie Constance Corsi, Damien Coyle, F De Vico Fallani, S Enriquez-Geppert, P Figueiredo, et al. A user-centred approach to unlock the potential of non-invasive bcis: an unprecedented international translational effort. 2020.

[225] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society*, 67(2):301–320, 2005.