Université du Québec
Institut national de la recherche scientifique
Centre Énergie Matériaux Télécommunications

Hybrid Brain-Computer Interfaces: Improving Mental Task Classification Performance through Fusion of Neurophysiological Modalities

Par

Hubert J. Banville

Mémoire présenté pour l'obtention du grade de
Maître es Sciences, M.Sc.
en télécommunications

Novembre 2015

Jury d'évaluation

Examinateur interne et président du jury	Prof. Douglas O'Shaughnessy INRS-ÉMT
Examinateur externe	Prof. Sarah D. Power Memorial University of Newfoundland
Directeur de recherche	Prof. Tiago H. Falk INRS-ÉMT

Acknowledgements

I would like to thank Dr. Tiago H. Falk for his guidance and mentorship in completing this project, as well as in many other investigations carried out towards the completion of my thesis. Many thanks to the MuSAE Lab members who helped me learn so much over the course of this project, especially Rishabh Gupta and Raymundo Cassani, whose analytic and practical skills were indispensable, and who helped me collect data. I would also like to express my gratitude to the subjects who participated in this project. The last great thanks goes to my life partner Colleen Gillon, for all her love, support and invaluable input throughout this project.

I acknowledge the financial support of the Fonds de recherche québécois sur la nature et les technologies.

Résumé

Ce projet démontre l'utilisation de multiples techniques de neuroimagerie pour la classification d'états mentaux dans le contexte d'une interface cerveau-ordinateur (ICO). Basé sur de récentes études en électroencéphalographie (EEG) et en spectroscopie proche-infrarouge (SPIR) ayant démontré que des tâches telles que l'imagerie motrice et l'arithmétique mentale induisent des patrons d'activité neurophysiologique particuliers, un paradigme d'ICO hybride (ICOh) combinant l'EEG et la SPIR est développé afin d'améliorer la performance de classification binaire d'un tel système.

Des signaux d'EEG et de SPIR ont été enregistrés sur neuf participants réalisant sept tâches mentales (génération de mots, rotation mentale, soustraction mentale, chant mental, navigation mentale, imagerie motrice et imagerie de visages). Ensuite, des machines à vecteurs de support avec régularisation l1 ont été entraînées sur chacune des paires de tâches possibles 1) en EEG, 2) en SPIR, et 3) en EEG et en SPIR conjointement, afin d'identifier les meilleures tâches et d'évaluer l'utilité d'une approche multimodale. L'évaluation subjective de la charge mentale induite par chaque tâche a aussi été collectée afin d'en estimer la facilité d'utilisation et d'identifier des traits caractéristiques neurophysiologiques qui permettraient de prédire l'expérience subjective d'un utilisateur à travers une analyse des corrélations entre traits caractéristiques et dimensions subjectives.

L'approche multimodale EEG-SPIR a permis d'améliorer la valeur maximale du kappa de Cohen (mesure de l'accord entre deux observateurs) de 0.03 lorsque basée sur des fenêtres d'une seconde, ce qui correspond à une augmentation de l'exactitude de classification de 1.5% pour des classes équilibrées. Cette augmentation s'est révélée encore plus importante (variation de 0.17, soit 8.5% en exactitude) lorsque les cinq dernières secondes des tâches étaient prises en compte. L'analyse de l'EEG et de la SPIR a de plus permis de révéler l'importance relative de différentes régions du cerveau et de types de traits caractéristiques, alors que l'analyse des corrélations a permis d'identifier les traits caractéristiques les plus prometteurs pour prédire l'expérience subjective d'un utilisateur, bien qu'une évaluation plus approfondie sera nécessaire pour valider ces résultats. Ces travaux fournissent une base solide pour d'autres études sur les ICOh cherchant à améliorer la performance de classification vers la réalisation d'ICO plus efficaces et flexibles.

Mots-clés interfaces cerveau-ordinateur, hybride, multimodal, EEG, SPIR, tâches mentales, neuroimagerie fonctionnelle

Abstract

In this work, we demonstrate the use of multiple neuroimaging modalities to classify brain states in a Brain-Computer Interface. Based on recent electroencephalography (EEG) and near-infrared spectroscopy (NIRS) studies that showed that tasks such as motor imagery and mental arithmetic induce specific neural response patterns, we propose a hybrid BCI (hBCI) paradigm in which EEG and NIRS data are fused to improve binary classification performance.

We recorded full-head EEG and partial-head NIRS data from nine participants performing seven mental tasks (word generation, mental rotation, subtraction, singing and navigation, motor and face imagery). l1-Support Vector Machines were trained for each possible pair of tasks on 1) EEG features, 2) NIRS features, and 3) EEG and NIRS features together, to identify the best pairs and assess the usefulness of a multimodal approach. Subjective ratings were also collected to assess the usability of each task and carry out a correlation analysis to discover features that could help predict the subjective experience of a user.

The EEG-NIRS multimodal approach led to an average increase in peak kappa (an inter-rater agreement metric) of 0.03 when using features extracted from one-second windows (equivalent to an increase of 1.5% in classification accuracy for balanced classes). The increase was much stronger (change of 0.17, corresponding to an 8.5% accuracy increase) when focusing on the last five seconds of the 15-second trials. The EEG and NIRS analyses further unveiled relevant brain regions and important feature types, while preliminary results from the correlation analysis uncovered features that could be used to predict the subjective experience of a BCI user. This work provides a basis for future hBCI studies aiming to improve classification performance toward more efficient and flexible BCIs.

Keywords brain–computer interfaces, hybrid, multimodal, EEG, NIRS, mental tasks, functional neuroimagery

Sommaire récapitulatif

0.1 Introduction

Une interface cerveau-ordinateur (ICO) est un système de communication qui relie le cerveau à un ordinateur sans faire usage des voies de communication naturelles du cerveau, telles que les mouvements musculaires [1]. Ces systèmes permettent de remplacer, restaurer, améliorer ou compléter les sorties naturelles du système nerveux central [2], et ont été utilisés pour des applications cliniques, de même que pour le divertissement et en éducation [3]. Plus particulièrement, un grand nombre d'études a porté sur l'application de ces systèmes comme technologie d'assistance, par exemple pour des individus souffrant d'une maladie dégénérative (telle que la sclérose latérale amyotrophique) ou causée par un accident menant à une paralysie [4]. En effet, à cause de leur état de santé, ces individus ne peuvent souvent plus communiquer ou interagir avec leur environnement, et une ICO devient alors leur unique moyen de communication, menant à une amélioration dramatique de leur qualité de vie.

Afin de reconnaître l'intention d'un utilisateur, la majorité des designs d'ICO font usage d'une technique de neuroimagerie fonctionnelle nommée électroencéphalographie (EEG). Cependant, d'autres techniques de neuroimagerie peuvent être utilisées pour fournir de l'information sur l'activité cérébrale. En effet, la magnétoencéphalographie (MEG), la spectroscopie proche-infrarouge fonctionnelle (SPIR), l'imagerie par résonance magnétique fonctionnelle (IRMf) [5–11], entre autres, ont déjà été utilisés à cet effet. Une approche récente nommée *interfaçage cerveau-ordinateur multimodale* consiste quant à elle à utiliser plus d'une modalité à la fois, incluant au moins une modalité neurophysiologique (mais pouvant aussi inclure des modalités non-neurophysiologiques) [12, 13], afin d'améliorer la performance du système par rapport à une ICO unimodale. Cette approche fait partie de la famille plus large d'ICO hybrides (ICOh), combinant différents patrons d'activité cérébrale et/ou différentes modalités afin d'atteindre ces effets désirables.

Une ICO se base sur la reconnaissance d'un ou de plusieurs patrons d'activité neurophysiologique qui peuvent être produits volontairement par l'utilisateur afin de permettre le contrôle d'un appareil par la pensée. Malgré leur utilisation fréquente dans la littérature, certains patrons spécifiques tels que la désynchronisation et la synchronisation liées à l'événement (DLÉ/SLÉ), les potentiels évoqués (PÉ) comme le P300 et les potentiels évoqués visuels stationnaires (PÉVS), n'offrent pas une performance optimale chez tous les utilisateurs d'ICO. La variabilité inter-sujet, un phénomène qui se manifeste par

des différences significatives entre les signaux neurophysiologiques de différents utilisateurs, signifie que certains patrons seront mieux adaptés à certains utilisateurs, tout en étant sous-optimaux chez d'autres [14]. Il est donc possible d'améliorer significativement la performance et la facilité d'utilisation d'une ICO en identifiant les patrons optimaux sur une base individuelle pour chaque utilisateur. Deuxièmement, les utilisateurs victimes d'un traumatisme cérébral peuvent souffrir de dommages à certaines régions du cerveau entraînant une perte de fonctions cérébrales. Pour ces utilisateurs cibles, il est alors nécessaire d'identifier des patrons différents qui recrutent d'autres régions non-affectées du cerveau.

Ce travail a pour but d'identifier et de valider la recherche existante sur les tâches mentales optimales qui peuvent être utilisées pour contrôler une ICO en EEG, en SPIR et en EEG-SPIR, et d'explorer l'utilité de ces modalités dans l'atteinte d'ICO plus performantes.

0.1.1 Objectifs de recherche et contribution

Afin d'évaluer le potentiel de différentes tâches mentales pour le contrôle d'une ICO, ce travail se concentre sur les deux questions suivantes:

1. Quelles tâches mentales pourraient être utilisées afin de contrôler une ICO avec une précision minimale de 70% (tel que recommandé par [15])?
2. Dans quel contexte une interface multimodale EEG-SPIR assure-t-elle une meilleure performance qu'une ICO basée uniquement sur l'EEG ou sur la SPIR? De plus, l'utilisation conjointe de ces modalités pourrait-elle fournir de l'information supplémentaire utile au développement d'ICOs plus faciles d'utilisation?

L'utilité de l'EEG et de la SPIR pour la classification de tâches mentales a déjà été démontrée (voir Section 0.2.2). La plupart des études sur le sujet ont pu identifier des paires de tâches mentales permettant d'atteindre 70% de précision de classification, toutefois peu d'entre elles ont évalué plus de trois tâches à la fois. De plus, rares sont les études ayant analysé le comportement des traits caractéristiques et des classifieurs utilisés, qui pourtant permettrait d'amasser des informations précieuses sur l'importance relative de chaque type de trait caractéristique et sur leur distribution spatiale [16]. De plus, seulement trois études [17–19] (décrites brièvement à la Section 0.2.2) ont étudié l'impact d'une fusion EEG-SPIR pour la classification de tâches mentales. Seulement trois types de tâches mentales ont été étudiés dans ces travaux (imagerie motrice, soustraction mentale et décompte mental), laissant ouverte la question de comment la fusion peut aider à améliorer la performance d'autres types de tâches mentales. Afin de combler ces lacunes, ce mémoire vise à

1. Comparer et analyser la performance (incluant l'appréciation subjective) de différentes paires de tâches mentales en utilisant l'EEG, la SPIR, et une approche multimodale EEG-SPIR.
2. Évaluer le pouvoir explicatif de différents traits caractéristiques neuroélectrophysiologiques et neurohémodynamiques en lien avec l'expérience subjective d'un utilisateur d'ICO.

0.1.2 Organisation du document

Ce travail est organisé comme suit. Dans cette section, la nécessité d'améliorer la performance des ICO par l'identification de nouvelles tâches mentales et l'utilisation de modalités additionnelles a été présentée, de même que les objectifs de recherche de ce travail. À la Section 0.2, la pertinence du projet de recherche est justifiée par un bref exposé du domaine de l'interfaçage humain-ordinateur et la présentation de l'état-de-l'art de la classification multimodale de tâches mentales, suivis d'une description des améliorations proposées par ce projet. La méthodologie est décrite à la Section 0.3, couvrant la collecte des signaux neurophysiologiques, le traitement et l'analyse des données. Ensuite, les résultats sont présentés (Section 0.4) puis discutés à la lumière de l'état-de-l'art (Section 0.5). Finalement, un résumé de ce travail est présenté et des directions futures de recherche sont proposées dans la conclusion de ce mémoire (Section 0.6).

0.2 Interfaces cerveau-ordinateur: progrès récents et défis actuels

0.2.1 Structure et définitions

En se basant sur le contournement des voies de sortie normales du cerveau [1], les ICO ont été initialement développées pour venir en aide aux individus paralysés ou souffrant de déficience motrice grave [20]. Différentes applications pour les utilisateurs en bonne santé ont également été proposées, par exemple pour les jeux vidéo [21, 22] et le monitorage affectif [23]. Malgré l'existence de plusieurs modalités de neuroimagerie fonctionnelle, les ICO basées sur l'EEG ont fait l'objet de la majorité des travaux dans ce domaine grâce à la bonne compréhension de ses caractéristiques de base, son accessibilité, sa portabilité et sa haute résolution temporelle [1, 6].

Généralement, une ICO se base sur la reconnaissance de patrons d'activité cérébrale distincts, ou *patrons de réponse neuronale*, qui peuvent être provoqués par l'utilisateur de manière interne (consciemment) ou inconsciemment, en utilisant des stimuli externes [24]. Récemment, une nouvelle approche nommée interfaçage humain-ordinateur hybride (ICOh) a émergé, cherchant à combiner deux ou plusieurs ICO, ou une ICO avec une autre modalité d'entrée telle que des biosignaux ou des dispositifs d'interface homme-machine classiques [13, 25]. Une ICOh permet de capitaliser sur les forces de plusieurs modalités afin de pallier leurs faiblesses respectives. Les ICOh sont prometteuses en raison de leur flexibilité accrue (permettant à l'interface de s'adapter à des utilisateurs différents), de leur robustesse (la combinaison des entrées peut faciliter la reconnaissance des intentions de l'utilisateur) et du nombre plus élevé de commandes qu'elles permettent.

Une ICO se compose généralement de plusieurs modules formant une boucle fermée entre l'utilisateur et l'ordinateur [24, 26], tel que présenté à la Figure 2.1. En premier lieu, une modalité d'imagerie cérébrale (EEG, SPIR, etc.) est utilisée pour recueillir des informations sur l'activité mentale de l'utilisateur. Pour une ICOh, des modalités supplémentaires peuvent être utilisées en plus de celles mentionnées ci-dessus:

par exemple, la photopléthysmographie, la réponse électrodermale, des interfaces humain-ordinateur traditionnelles (par exemple, une souris ou un clavier), etc. [13]. Nous décrivons maintenant plus en détail les deux modalités le plus souvent utilisées dans la littérature, et qui seront l'objet de ce travail : l'EEG et la SPIR.

Électroencéphalographie

Le cerveau est composé de millions de milliards (10^{11}) de cellules appelées neurones [27]. Ces cellules sont caractérisées par leur capacité unique à transmettre de l'information par l'intermédiaire de phénomènes électrochimiques. Dans le cortex, la couche périphérique du cerveau qui gère les fonctions supérieures, les populations de neurones sont hautement structurées. À cause de cette structure, lorsqu'activées de manière synchronisée, ces populations produisent un potentiel électrique d'environ 100 μV qui peut être mesuré sur le scalp [27]. Ce potentiel électrique est en fait la somme d'un grand nombre de champs électriques originaires de différentes parties du cerveau et atténués par différents tissus comme le cerveau lui-même, les méninges, le crâne et la peau. L'EEG offre ainsi une faible résolution spatiale : un signal d'EEG contient de l'information provenant de nombreuses régions du cerveau, qui sont souvent difficiles à démêler.

Pour enregistrer ces potentiels électriques, des électrodes sont positionnées directement sur le scalp (par exemple, suivant le système international 10-20 [27] (voir la Figure 2.2), souvent avec un gel ou pâte conducteurs, ou une solution saline pour réduire l'impédance entre les capteurs et la peau [28]. L'EEG est malgré tout très sensible à plusieurs sources de signaux indésirables appelés *artéfacts*, qui proviennent par exemple d'autres processus physiologiques (battements cardiaques, mouvements des yeux, mouvements musculaires ou même mouvements de la langue) ou du déplacement des capteurs par rapport à la peau.

Des informations importantes sur les fonctions cognitives peuvent être décodées à partir des composantes fréquentielles de l'EEG [29–31]. En effet, certaines bandes de fréquences sont souvent utilisées pour étudier différents états cérébraux : les bandes Δ (0-4 Hz), θ (4-8 Hz), α (8-12 Hz), β (12-30 Hz) et γ (30-100 Hz) sont par exemple chacune corrélées avec des états ou des processus mentaux spécifiques. D'autres informations plus complexes peuvent être extraites de l'EEG pour en étudier le comportement [24].

Spectroscopie proche-infrarouge fonctionnelle

Alors que l'EEG mesure directement l'activité neuronale via les potentiels électriques qu'elle génère, la SPIR repose sur les changements des propriétés optiques du sang pour mesurer indirectement l'activité neuronale [32]. Chaque fois qu'un groupe de neurones est activé, une quantité accrue d'oxygène est nécessaire dans les tissus avoisinants afin de supporter l'augmentation de la consommation d'énergie. Une plus grande quantité d'oxygène doit donc être transportée dans le sang des poumons au cerveau. Pour ce faire, le corps humain utilise l'hémoglobine, une molécule qui existe dans deux états: oxygéné

(HbO) et désoxygénée (HbR). Une molécule d'hémoglobine désoxygénée devient oxygénée lorsque quatre molécules d'oxygène s'y attachent. Par coïncidence, ce changement d'état a également un effet sur les propriétés optiques de la molécule d'hémoglobine, qui pour cette raison est qualifiée de *chromophore*. En effet, le spectre d'absorption de l'HbR diffère de celui de l'HbO, en particulier dans le spectre proche-infrarouge (700 à 900 nm). En mesurant l'absorption de photons réfléchis à travers un médium contenant de l'hémoglobine, il devient ainsi possible de mesurer les variations locales d'oxygénation du sang, et ainsi de mesurer indirectement l'activité neuronale.

Tel qu'illustré à la Figure 2.3, cette mesure est réalisée avec des sources et des détecteurs positionnés sur le scalp et séparés d'environ 3 cm. Les photons émis par la source pénètrent et sont reflétées par les tissus; certains d'entre eux se rendent ainsi jusqu'à la couche externe du cortex, puis sont reflétés jusqu'à un détecteur voisin (voir la Figure 2.3) [33]. L'équation reliant l'intensité de la lumière incidente I_λ et l'intensité de la lumière transmise $I_{0,\lambda}$, appelée *loi de Beer-Lambert modifiée* (MBLL), est ensuite utilisée pour déterminer les changements de concentration des chromophores (voir les Équations 2.1 et 2.2) [32].

Les changements de concentration en HbO et HbR (Δ[HbO] et Δ[HbR]) se développent sur une période d'environ six secondes, délai causé par les mécanismes reliant l'activité neuronale et la circulation sanguine, regroupés sous le terme de *réponse hémodynamique* [9] (voir la Figure 2.4).

À l'instar de l'EEG, la SPIR est contaminée par des artéfacts physiologiques, de mouvement, et externes [33]: les battements cardiaques, la respiration et les dérives en basses fréquences causées par l'instrumentation doivent ainsi être filtrés des signaux de SPIR. Toutefois, contrairement à l'EEG, les signaux mesurés avec la SPIR proviennent d'une région précise du cerveau (de l'ordre de quelques centimètres) [34].

Paradigmes d'utilisation d'ICOs

Les ICOs utilisant l'EEG peuvent se baser sur différents patrons de réponse neurale. L'**activité sensorimotrice** émerge de l'activité rythmique mesurée entre 8 à 12 Hz et 13 à 30 Hz (les rythmes μ et β) dans les cortex frontal et pariétal, lorsque ces régions du cerveau sont inactives. Cependant, chaque fois qu'un mouvement volontaire est exécuté ou imaginé, ces rythmes perdent de leur intensité, un phénomène nommé désynchronisation liée à l'événement (DLÉ). De même, une fois que le mouvement (ou son imagination) est terminé, ces rythmes réapparaissent, produisant une synchronisation liée à l'événement (SLÉ). L'apparition ou l'estompage de ces rythmes permet donc d'identifier une activité cérébrale consciente. Les **potentiels évoqués** (PÉ) sont quant à eux des déviations positives ou négatives observées dans le signal d'EEG, provoquées par des stimuli spécifiques. Le PÉ le plus souvent utilisé dans une ICO est le P300, une déviation positive dans l'activité électrique du cortex pariétal se produisant environ 300 ms après la présentation d'un stimulus rare, comme un stimulus visuel, auditif ou tactile. Les **potentiels évoqués visuels stationnaires** (PÉVS) sont provoqués par des stimuli oscillatoires présentant une composante de fréquence constante. En réponse à cette stimulation rythmique, le cortex

sensoriel produit une activité spectrale de même fréquence que le stimulus [35]. Finalement, différentes **tâches mentales** sont également utilisées pour susciter des patrons d'activité cérébrale spécifiques à chaque tâche. Par exemple, le calcul mental, la rotation mentale, l'attention spatiale et la sensation sélective produisent tous des patrons d'activation qui recrutent différentes régions du cerveau et évoluent différemment [36–38].

D'autre part, les patrons de réponse neuronale utilisés avec des modalités neurophysiologiques basées sur l'activité hémodynamique du cerveau sont généralement moins diversifiés. Par exemple, en SPIR et en IRMf, une augmentation ou diminution localisée du débit sanguin et de l'oxygénation peut être mesurée afin de détecter l'activation de certaines régions du cerveau [32]. La plupart des ICO hémodynamiques visent ainsi à classifier des états mentaux qui sont de nature à induire différents patrons d'activation spatio-temporelle [11].

Des signaux neurophysiologiques à l'ordinateur

Le traitement des signaux neurophysiologiques enregistrés est effectué selon trois étapes générales dans une ICO : 1) traitement des artéfacts, 2) génération et 3) transposition de traits caractéristiques. Lors de la conception d'une ICOh, l'intégration des différents signaux et modalités peut être faite à l'une des deux dernières étapes (voir la section 2.1.7).

Lors du traitement des artéfacts, les composantes indésirables qui découlent d'autres processus physiologiques ou d'événements physiques (comme le mouvement relatif entre l'utilisateur et les senseurs) et qui polluent le signal d'intérêt sont filtrés des signaux neurophysiologiques [39]. Le lecteur intéressé est référé à [40] et [41] pour une revue des techniques de traitement d'artéfacts en EEG et en SPIR, respectivement.

La génération de traits caractéristiques vise à extraire des caractéristiques pertinentes à l'analyse de l'activité physiologique à partir des signaux bruts [24]. Tout d'abord, les signaux sont préretraités afin d'isoler certaines de leurs composantes, telles qu'une gamme de fréquences ou une distribution spatiale, généralement avec différents types de filtres. Ensuite, des informations précises sont extraites des signaux prétraités, appelées *traits caractéristiques*, qui indiquent de manière plus ou moins fiable la présence ou l'absence d'une activité spécifique dans le cerveau de l'utilisateur (voir la Section 0.2.1). Enfin, avec les systèmes multimodaux, la quantité de données qui est générée peut être considérable, en particulier si plusieurs canaux sont utilisés par modalité, et si plusieurs traits caractéristiques sont extraits par canal. Il est donc nécessaire de sélectionner les plus informatifs d'entre eux, ce qui peut être fait en utilisant des algorithmes dédiés.

La transposition des traits caractéristiques est la dernière étape du traitement, et se sert des traits caractéristiques précédemment extraits et d'un algorithme de reconnaissance de formes tel qu'un classifieur ou un ensemble de règles prédéfinies pour prendre une décision sur l'état de l'utilisateur [42]. Selon la configuration du système, plusieurs de ces classifieurs peuvent fonctionner en parallèle, auquel cas leurs décisions respectives sont combinées et fusionnées par un ensemble de règles ou un métaclassifieur.

L'objectif de la classification est d'apprendre une fonction f_θ qui peut associer une nouvelle mesure x' à la classe y' à laquelle elle appartient. Dans le contexte d'une ICO, un classifieur doit identifier l'état mental d'un individu à partir de d traits caractéristiques reflétant ses processus neurophysiologiques. De nombreux algorithmes peuvent être utilisés à cette fin (voir [42] pour une excellente revue du domaine).

La plupart des ICOh mentionnées dans la littérature utilisent les mêmes classifieurs linéaires [14]. Ces classifieurs linéaires tentent de trouver l'hyperplan de dimension d qui sépare le mieux des points de données n dans leurs classes respectives y. Par exemple, l'algorithme des machines à vecteurs de support (SVM), basé sur une vue géométrique de la frontière de décision [43], cherche un hyperplan qui maximise la distance entre l'hyperplan lui-même et les points de données les plus proches - une distance appelée *marge* - tout en séparant les points de données selon leur classe respective (voir la Figure 2.5). Le problème d'optimisation correspondant est présenté à l'Équation 2.5.

Enfin, une fois l'état mental de l'utilisateur identifiée par un classifieur ou un autre algorithme, la décision est envoyée à un dispositif de commande, tel qu'un affichage, une prothèse, un système robotisé, etc. Ce dispositif permet normalement une rétroaction visuelle, auditive ou tactile, ou dans d'autres cas qui n'est simplement pas perceptible par l'utilisateur.

Les stratégies d'intégration de signaux multimodaux peuvent être classées selon le niveau auquel se fait cette fusion: au niveau des données, des traits caractéristiques, ou des décisions [44]. La fusion au niveau des données est utilisée lorsque des signaux sont fusionnés avant l'extraction de traits caractéristiques, évitant ainsi la perte d'information lors de la fusion, mais augmentant en même temps la sensibilité au bruit et à la défaillance d'une des modalités. La fusion au niveau des traits caractéristiques est utilisée lorsque les signaux d'origine sont traités individuellement et leurs traits caractéristiques extraits avant d'être fusionnés. Cette méthode diminue la quantité d'information disponible lors de la fusion, mais améliore la robustesse du système. Finalement, la fusion au niveau des décisions est le plus haut niveau auquel les signaux peuvent être intégrés. Typiquement, cela implique que les différents signaux soient traités individuellement et classifiés (produisant ainsi une décision pour chaque signal) avant d'être fusionnés.

0.2.2 Classification de tâches mentales avec des ICO unimodales et multimodales

Diverses tâches mentales recrutant différentes parties du cerveau, telles que la soustraction et la rotation mentale, ont été récemment étudiées en EEG [30, 36, 45–53], en SPIR [54–62], en imagerie Doppler transcranienne (DTC) [63, 64], en SPIR-DTC [65], et en IRMf [66]. Les tâches mentales les plus fréquemment utilisées dans ces articles sont la soustraction mentale, la rotation mentale, diverses tâches de fluidité verbale, l'imagerie motrice et l'imagerie auditive. Ces études ont exploré la classification binaire ou multi-classe de tâches mentales, entre tâches ou entre un état de repos et une tâche, et dans la plupart des cas visaient à identifier la combinaison de tâches qui atteindrait la plus haute performance dans une ICO.

Études unimodales en EEG et SPIR

Dans les études d'EEG qui ont évalué plus de trois tâches différentes, des combinaisons de tâches casse-tête, c'est-à-dire une tâche qui implique la résolution d'un problème induisant un certain degré de charge mentale, et de tâches d'imagerie dynamique ont systématiquement engendré les plus hautes performances. Dans [47], la classification binaire du chant et du calcul mental (addition ou soustraction) a atteint le top cinq des meilleures combinaisons de tâches dans quatre des cinq participants, engendrant des valeurs supérieures à 93 % de précision de classification. Bien que visant à différencier une tâche mentale du repos au lieu d'une autre tâche mentale, Faradji *et al.* ont rapporté qu'une tâche de multiplication mentale pourrait produire un taux de vrais positifs supérieur à 70% tout en maintenant un taux de faux négatifs nul [48]. Friedrich et ses collaborateurs ont plus d'une fois soutenu l'optimalité d'une combinaison de tâches casse-tête et de tâches d'imagerie dynamique [36, 51, 53]. Dans [36], des paires formées à partir de sept tâches mentales différentes ont été évaluées sur neuf participants. L'association de mots, la soustraction mentale, la rotation mentale et l'imagerie motrice ont été identifiées comme les tâches les plus discriminantes, conduisant à un κ de Cohen (mesure de l'accord entre observateurs) supérieur à 0,7. L'étude des patrons de DLÉ/SLÉ évoqués par ces tâches a révélé des caractéristiques particulières qui pourraient expliquer pourquoi ce type de combinaison est optimal [36]. En outre, les auteurs ont mesuré l'appréciation subjective de chaque tâche: malgré une forte variabilité et l'absence de différences significatives entre les tâches, la génération de mot a reçu la meilleure évaluation tandis que la soustraction mentale a reçu la pire. De même, dans [51], une combinaison de soustraction mentale et d'imagerie motrice a mené à une performance inter-session systématiquement élevée chez sept des neuf participants (κ supérieur à 0,6). Enfin, dans [53], huit participants ont appris à contrôler d'une ICO à quatre classes dans laquelle des tâches mentales avaient été choisies sur une base individuelle. Les combinaisons de tâches les plus fréquemment sélectionnées incluaient l'imagerie motrice, la rotation mentale et autres tâches casse-tête, conduisant à des performances moyennes variant entre 44 et 84% de précision de classification.

Seules deux études ont tenté de classifier plus de deux tâches à la fois en utilisant la SPIR. Dans [59], l'utilisation de l'information hémodynamique du cortex préfrontal a permis d'atteindre une précision de classification binaire d'environ 60%, obtenue avec des paires formées de la soustraction mentale, de la génération de mot et de la rotation mentale. Une évaluation plus complète des tâches mentales utilisant la SPIR a été rapportée dans [6], profitant d'un montage couvrant l'ensemble de la tête. Une précision de classification d'environ 71% a ainsi été obtenue pour des combinaisons d'imagerie motrice, de multiplication et de rotation mentale chez sept participants.

Études multimodales en EEG-SPIR

Certaines études ont exploré l'effet de l'ajout de modalités supplémentaires à la classification de différentes tâches mentales, afin d'améliorer le nombre de tâches pouvant être classifiées ou d'améliorer la robustesse d'une tâche précise. Des combinaisons de SPIR et d'EEG pour reconnaître les tâches d'image-

rie ou d'exécution motrice ont été étudiées suivant trois approches différentes. Fazli *et al.* ont utilisé une procédure basée sur la fusion de décisions avec des classifieurs individuels pour l'EEG et les niveaux de concentration d'HbO et d'HbR, afin de distinguer un mouvement imaginé ou exécuté de la main gauche et de la main droite [17]. Avec cette approche, une augmentation moyenne de 5% de précision a été obtenue chez 13 des 14 participants pour la tâche d'imagerie motrice, générant une précision moyenne de 83,2%. Les auteurs ont reconnu les inconvénients du délai introduit par la réponse hémodynamique, rendant difficile l'utilisation de la SPIR au sein d'ICO pratiques. Une approche différente a été explorée par Khan *et al.* afin d'augmenter le nombre de commandes d'entrée disponibles pour un système de navigation [18]. Alors que les mouvements vers la gauche et la droite étaient contrôlés avec l'imagerie motrice mesurée en EEG, les mouvements vers l'avant et l'arrière étaient contrôlés par une tâche de décompte mental ou de soustraction mentale mesurées en SPIR. Des précisions de classification de 94,7% et de 80% en moyenne ont été obtenues pour les tâches en EEG et en SPIR, respectivement. Enfin, dans [19], la SPIR a été utilisée pour détecter l'exécution d'une tâche d'imagerie motrice, avant de déclencher la classification de sa latéralité entre mouvement de la main gauche et de la main droite en EEG. Cette approche a conduit à un faible taux de faux positifs de 7% et un taux de vrais positifs de 88%.

Plusieurs autres résultats concernant l'état-de-l'art en interfaçage humain-ordinateur hybride sont disponibles dans une revue de la littérature en voie de publication [14].

0.2.3 Améliorations suggérées

Ce mémoire vise l'amélioration des connaissances sur de nouvelles tâches de contrôle d'ICO, en adoptant une approche hybride dans laquelle la SPIR et l'EEG sont enregistrés simultanément. Cette approche complète celles d'études antérieures comme [36, 61] en analysant précisément le fonctionnement d'un tel système hybride et en évaluant comment elle peut fournir un gain de performance par rapport aux approches unimodales standards. L'impact des fenêtres d'extraction de traits caractéristiques, de même que l'évolution dynamique de la performance de classification, sont également explorés pour mieux comprendre comment les ICOh peuvent être améliorées. Enfin, une analyse exploratoire est réalisée afin d'identifier les caractéristiques neurophysiologiques qui pourraient être utilisées pour prédire l'expérience subjective d'un utilisateur d'ICO, menant potentiellement à des ICOh conscientes de l'état de leur utilisateur.

0.3 Méthodologie

Douze participants (cinq femmes, trois gauchers, âgés de 20 à 31 ans, moyenne de 24,6 ans), parlant couramment anglais et/ou français, ont participé à notre étude EEG-SPIR. Les participants devaient compléter trois sessions d'enregistrement de deux à trois heures chacune pour lesquelles ils ont reçu une rémunération, à l'intérieur d'une période de trois à cinq semaines. Les participants ont déclaré ne pas avoir de troubles neurologiques et n'avoir aucune expérience préalable avec des ICO. Chaque participant a

accepté les termes et conditions de l'étude, approuvée par le comité d'éthique de l'université. Les données de deux participants ont été rejetées car trop bruitées; celles d'un troisième participant ont été rejetées parce que ce participant n'avait complété qu'une seule des trois sessions requises. Par conséquent, les données provenant d'un total de neuf participants ont été utilisées dans cette étude.

Les participants devaient effectuer sept types de tâches mentales suscitant des patrons de réponse neurale spécifiques, tel qu'identifiées dans des études précédentes utilisant l'EEG et la SPIR [36, 61].

Rotation mentale (ROT) Deux solides en forme de L, identiques ou un étant la réflexion de l'autre, mais dans chaque cas dans un état différent de rotation, ont été présentés aux participants. Les participants devaient imaginer faire tourner l'un des deux solides afin d'identifier s'il s'agissait de deux solides identiques ou d'images miroir.

Génération de mots (WORD) Les participants devaient trouver autant de mots que possible commençant par une lettre choisie au hasard présentée sur l'écran. Des mots en anglais ou en français étaient requis selon la langue choisie par le participant au début de l'expérience.

Soustraction mentale (SUB) Les participants devaient soustraire successivement deux nombres à un ou deux chiffres d'un nombre à trois chiffres (par exemple 214-9 = 205, 205-13 = 192, etc.).

Chant mental (SING) Les participants devaient imaginer chanter une chanson choisie à l'avance, si possible avec les paroles, tout en se concentrant sur la réponse émotionnelle qu'elle suscitait chez eux.

Navigation mentale (NAV) Les participants devaient imaginer marcher d'un endroit à un autre dans leur résidence actuelle ou précédente, tout en se concentrant sur leur orientation spatiale (par exemple, marcher de leur chambre au réfrigérateur).

Imagerie motrice (MI) Les participants devaient imaginer tapoter les doigts de leur main droite.

Imagerie de visage (FACE) Les participants devaient imaginer le visage d'un ami, tel que mémorisé à partir d'une photo apportée à la séance d'enregistrement.

D'après la description des types de tâches faite par Friedrich *et al.* [36], la génération de mots et la soustraction mentale sont des tâches *casse-tête*, car elles nécessitent la résolution d'un problème; le chant mental, la navigation mentale et l'imagerie motrice sont des tâches d'imagerie dynamiques; et l'imagerie de visage est une tâche d'imagerie statique. La rotation mentale est ici classée parmi les tâches casse-tête plutôt que les tâches d'imagerie dynamique car les participants devaient donner une réponse (soit si les solides sont identiques ou non). Cette modification a été motivée par la littérature sur la rotation mentale, qui utilise souvent cette version de la tâche [67, 68].

Avant le début de l'enregistrement, les participants ont réalisé chaque tâche ouvertement afin de s'assurer que les instructions étaient bien comprises. Le déroulement d'une répétition d'une tâche mentale est illustré à la Figure 3.1. Une session est composée de quatre sous-sessions, dans lesquelles chaque type

de tâche mentale est répété quatre fois, ce qui donne un total de 28 exécutions de tâche par sous-session. Chaque sous-session commence et finit par une période de repos de 30 secondes, pendant laquelle les participants doivent rester dans un état mental neutre et fixer la croix au centre de l'écran. Avant chaque exécution de tâche, un compte à rebours indique la prochaine tâche à effectuer à l'aide du pictogramme associé (voir Figure 3.1). Une fois le compte à rebours terminé, les participants doivent exécuter la tâche mentale spécifiée pendant 15 secondes. Chaque exécution est ensuite suivie d'une période de repos d'une durée aléatoire variant entre 10 et 15 s. Une fois une sous-session terminée, les participants pouvaient s'étirer légèrement, boire ou manger une collation avant de reprendre l'enregistrement. Les stimuli et le questionnaire ont été générés avec le logiciel Presentation (Neural Behavioral Systems, É.U.).

Les données d'EEG ont été enregistrées avec un système ActiveTwo (Biosemi B.V., Amsterdam, Pays-Bas) de 62 sondes (plus deux mastoïdes et CMS et DRL) et quatre électrodes d'électrooculogramme (EOG). Le système international 10-10 a été utilisée pour le placement des électrodes, sans AF7 et AF8 toutefois, dont les emplacements ont été utilisés pour des sondes de SPIR (voir Figure 3.2). Tous les signaux ont été numérisés à une fréquence d'échantillonnage de 512 Hz.

Les données de SPIR ont été enregistrées avec un système NIRScout (NIRx Medical Technologies, Los Angeles, É.U.), avec 16 sources (longueurs d'onde de 760 et 850 nm) et 24 détecteurs. Les optodes ont été placées sur le même casque que les électrodes d'EEG. Les lobes frontaux, centraux, temporaux et pariétaux ont été ciblés par le montage utilisé, suivant le système étendu 10-5 (voir la Figure 3.2). La montage n'a pas été étendu au lobe occipital en raison de la plus faible qualité des signaux de SPIR dans cette région et du nombre restreint d'optodes disponibles. Un total de 60 canaux de SPIR échantillonnés à 4,46 Hz ont été obtenus en utilisant les paires source-détecteur séparées par environ 3 cm.

À chaque session d'enregistrement, les participants ont été invités à remplir un questionnaire d'évaluation de leur expérience subjective des tâches mentales exécutées. Les questions étaient basées sur la première partie du test NASA TLX [69], dont la version française a été obtenue dans [70]. En outre, les participants ont été invités à classer les tâches mentales par ordre de préférence. Le Tableau 3.1 fait état des différents items mesurés par ce test.

0.3.1 Analyse des signaux neurophysiologiques

EEG

Les données brutes d'EEG ont été prétraitées à l'aide du logiciel EEGLAB [71]. Les signaux ont été importés et référencés à Cz avant d'être rééchantillonnés à 256 Hz, puis filtrées entre 0,5 et 100 Hz et à 60 Hz par un filtre coupe-bande. Les canaux trop bruités ont été identifiés visuellement et rejetés pour chaque session individuellement. Les exécutions de tâches contaminées par des artéfacts intenses ont également été rejetées. Un nettoyage supplémentaire des signaux a été réalisé par une méthode semi-automatique [72]. Les canaux rejetés ont été réinterpolés et tous les canaux ont été reréférencés à la moyenne de toutes les électrodes. Enfin, les signaux de chaque période d'exécution ont été normalisés

en utilisant la période de 300 ms avant le début de la tâche. De plus, afin de permettre une analyse physiologique des données, la méthode de la variance inter-essai proposée dans [73] a été utilisée pour calculer les valeurs de DLÉ/SLÉ pour chaque tâche mentale.

Les expériences de classification menées dans ce travail ont porté sur les bandes de puissance classiques uniquement. Cependant, des traits caractéristiques supplémentaires liés à des processus oscillatoires, dont l'utilité à été démontrée dans d'autres études d'EEG [74, 75], ont également été calculés, mais seulement utilisés dans l'analyse des corrélations (voir Section 0.3.3). Par conséquent, cinq types de traits caractéristiques différents ont été extraits pour chaque électrode, donnant un total de 1894 traits caractéristiques:

- Log-puissance dans différentes bandes de fréquences et ratios: δ (0-4 Hz), θ (4-8 Hz), $\alpha_{inférieur}$ (8-10 Hz), $\alpha_{supérieur}$ (10-12 Hz), $\beta_{inférieur}$ (12-21 Hz), $\beta_{supérieur}$ (21-30 Hz), $\gamma_{inférieur}$ (30-60 Hz), $\gamma_{supérieur}$ (60-80 Hz), θ à β (4-30 Hz), et spectre entier (0.1-100 Hz), de même que les ratios $\frac{\alpha_{total}}{\beta_{total}}$, $\frac{\theta}{\beta_{total}}$, et $\frac{\gamma_{supérieur}}{\beta_{total}}$.
- Modulation d'amplitude dans différentes bandes de fréquences [74, 75]: δ m-δ, θ m-δ, θ m-θ, α m-δ, α m-θ, β m-δ, β m-θ, β m-α, et β m-β.
- Cohérence d'amplitude et de phase pour des paires d'électrodes inter-hémisphériques [74, 76, 77]: MSC et Φ_{coh} sur 26 paires d'électrodes inter-hémisphériques (FP1-FP2 AF3-AF4, F7-F8, F5-F6, F3-F4, F1-F2, FT7-FT8, FC5-FC6, FC3-FC4, FC1-FC2, T7-T8, C5-C6, C3-C4, C1-C2, TP7- TP8, CP5-CP6, CP3-CP4, CP1-CP2, P9-P10, P7-P8, P5-P6, P3-P4, P1-P2, PO7-PO8, PO3-PO4, O1-O2) pour chaque bande de fréquence définie précédemment.
- Synchronisation globale de phase [78].

La nomenclature utilisée pour décrire ces traits caractéristiques est présentée en Annexe B. Chaque exécution de tâche mentale est subdivisée en multiples fenêtres de 1, 2, 3, 5 ou 15 secondes sans chevauchement, ou en fenêtres de 7,5 ou 10 secondes avec chevauchement, utilisées pour extraire les caractéristiques mentionnées ci-dessus (voir le Tableau 3.2).

SPIR

Le prétraitement des données de SPIR a été réalisé avec le logiciel HOMER2 [79]. Les canaux ayant un faible rapport signal-sur-bruit ont été identifiés en corrélant les densités optiques (filtrées entre 0,8 et 1,2 Hz) avec les canaux S15-D1 et S15-D2 sur la gauche et les canaux S16-D13 et S16-D14 sur la droite, avec chaque canal du côté gauche ou droit, respectivement. Ces 4 canaux frontaux étant les plus courts du montage utilisé, et n'étant généralement pas obstrués par des cheveux, une pulsation cardiaque claire peut y être observée entre 45 et 100 BPM, ce qui correspond à une fréquence cardiaque au repos normale. Les canaux positivement corrélés avec ces canaux de courte distance devraient donc être de bonne qualité, car ils contiennent assurément de l'information physiologique. Les canaux ont donc été

rejetés si la valeur-p médiane de la corrélation à travers toutes les époques d'une session était au-dessus de 0,05. Ensuite, la densité optique des canaux restants a été convertie en changements de concentration de HbO, HbR et HbT par la loi de Beer-Lambert modifiée. Enfin, 13 canaux artificiels ont été calculées en moyennant l'amplitude des canaux contenus dans les régions latéro-frontale, centro-frontale, temporale, temporo-pariétale, centrale et préfrontale (identifiés à la Figure 3.2). Ces mesures ont produit 73 canaux, chacun mesurant Δ[HbO], Δ[HbR] et Δ[HbT], donnant un total de 219 canaux de SPIR. Les réponses moyennes en Δ[HbO], Δ[HbR] et Δ[HbT] ont également été visualisées sur des cartes topographiques pour révéler les patrons d'activation hémodynamique de chaque tâche mentale.

Dans cette étude, les traits caractéristiques suivants ont été extraits en utilisant les mêmes fenêtres que pour l'EEG: l'amplitude moyenne, le délai (en secondes) d'atteinte de l'amplitude maximale et minimale de la fenêtre, la différence entre l'amplitude moyenne de la première moitié et de la seconde moitié de la moyenne de fenêtre, et les coefficients de régressions linéaire, quadratique et cubique modélisant les valeurs d'amplitude dans la fenêtre. La nomenclature utilisée pour décrire ces caractéristiques est présentée en Annexe B. Cela a donné un maximum de 2849 traits caractéristiques de SPIR.

0.3.2 Classification de l'EEG, de la SPIR et de la fusion EEG-SPIR

Des sous-ensembles de traits caractéristiques sélectionnés à partir des jeux de données d'EEG et de SPIR ont ensuite été utilisés pour entraîner un modèle de classification binaire sur chacune des paires de tâches. Pour l'EEG, les bandes de log-puissance ont été sélectionnés (558 fonctions), alors que pour la SPIR, les valeurs d'amplitude moyenne de Δ[HbO], Δ[HbR] et Δ[HbT] ont été utilisées (219 caractéristiques). La présélection d'un sous-ensemble de traits caractéristiques permet d'atténuer les problèmes causés par la malédiction de la dimensionalité. Tout d'abord, la procédure d'entraînement a été appliquée à l'EEG et la SPIR séparément pour évaluer la performance individuelle de chaque modalité. Ensuite, la même procédure a été appliquée sur un ensemble de données combinant les traits caractéristiques de l'EEG et la SPIR pour en évaluer l'impact sur la performance de classification.

Des SVM linéaires avec régularisation l_1 ont été entraînées sur ces jeux de données. La régularisation l_1 permet la sélection d'un faible nombre de traits caractéristiques importants [80, 81] et permet de minimiser les effets néfastes d'un jeu de données de haute dimensionalité mais avec relativement peu d'exemples. Une valeur par défaut de $2,0$ a été utilisé pour l'hyperparamètre C qui contrôle l'équilibre entre la perte la pénalisation l_1 des poids.

Pour chaque participant, 21 jeux de données ont été formés contenant chacun les exécutions de deux types de tâches différentes. Ces derniers ont ensuite été divisées en 10 partitions contenant chacune un nombre égal de répétitions pour chaque type de tâche. De plus, chaque type de fenêtre temporelle a été analysé indépendamment, c'est-à-dire qu'un classifieur unique a été entraîné pour chaque fenêtre de 1, 2, 5, 7.5, 10 et 15 secondes extraites.

Afin de minimiser autant que possible la latence d'une éventuelle ICO et permettre des taux élevés de transfert de l'information, l'essentiel de notre analyse se concentre sur les modèles entraînés sur

des traits caractéristiques provenant de fenêtres d'une seconde. Cette approche assure un délai minime entre la réalisation de la tâche mentale et la production d'une commande. En outre, les fenêtres d'une seconde permettent une analyse plus fine de l'évolution temporelle des traits caractéristiques sélectionnés et de la performance de classification, ce qui peut révéler des phénomènes physiologiques intéressants. Nous fournissons malgré tout une analyse de l'effet de la longueur des fenêtres sur les performances de classification.

La performance des classifieurs a été évaluée avec le κ de Cohen, qui mesure l'accord entre deux évaluateurs qui tentent de classifier N exemples en catégories exclusives (voir Équation 3.7) [36]. Le κ de Cohen est utile dans ce contexte car il tient compte du déséquilibre des classes produit par le rejet de certaines réalisations de tâches. Dans le cas d'un problème parfaitement équilibré, des valeurs de κ de 0 et de 1 sont équivalentes à 50% et 100% de précision de classification, respectivement. Notre analyse se concentre sur trois gammes de valeurs de κ: 0,0 à 0,4 (faible rendement), 0,4-0,7 (performance satisfaisante), et 0,7-1,0 (haute performance).

Importance des traits caractéristiques

L'importance des différents traits caractéristiques peut être analysée en utilisant les poids correspondants des SVM [82]. Comme les caractéristiques sont normalisées avant l'entraînement, les poids du modèle linéaire représentent l'importance de chaque trait caractéristique. Ainsi, un trait caractéristique assorti d'un poids absolu élevé peut donc être considéré comme étant plus important qu'un autre assorti d'un poids proche de 0. La valeur absolue des poids de chaque trait caractéristique a ainsi été moyennée puis comparée aux autres afin d'identifier l'information importance dont dispose le classifieur.

0.3.3 Analyse exploratoire des corrélations entre traits caractéristiques et évaluations subjectives

Les corrélations entre les traits caractéristiques extraits précédemment et les dimensions subjectives mesurées avec le questionnaire NASA TLX ont été mesurées. Cette analyse a été effectuée sur les traits caractéristiques provenant de fenêtres de 15 secondes, qui fournissent plus d'informations que les fenêtres plus courtes (en particulier pour la SPIR, compte tenu du délai de la réponse hémodynamique) et qui demeureraient utiles comme les applications de monitorage affectif ne nécessitent pas de taux de transfert d'information élevés.

Tout d'abord, la moyenne d'un trait caractéristique pour une tâche spécifique a été jumelée aux évaluations des cinq dimensions subjectives (exigence mentale, pression temporelle, performance, effort, frustration) données à chaque session par chaque participant. Ensuite, le coefficient de corrélation de Spearman (ρ_{Spear}) a été calculé pour chaque paire de trait caractéristique et de dimension subjective. Suivant une correction de Bonferroni, les corrélations absolues et les valeurs-p ont été seuillées à 0,4

et 0,05, respectivement, afin d'identifier les traits caractéristiques significativement corrélés avec les évaluations subjectives.

0.4 Résultats expérimentaux

Les données des participants S02 et S08 ont été rejetées dû à leur faible nombre de répétitions de bonne qualité et à leur somnolence observée chez ces participants pendant les enregistrements (voir les Tableaux 4.1 et 4.2), tandis que les données du participant S12 ont été rejetées car une seule session sur les trois requises avait été enregistrée. L'analyse qui suit se base donc sur les données de neuf participants avec au moins 53 canaux d'EEG valides, 27 canaux de SPIR valides, et 274 répétitions valides (81,55% des répétitions enregistrées).

0.4.1 Classification de tâches mentales sur une base individuelle

EEG

La performance de classification maximale des classifieurs entraînés sur des fenêtres d'une seconde d'EEG, pour chaque participant et paire de tâches, est présentée au Tableau 4.3. Tous les participants ont atteint des niveaux de performance satisfaisant et élevé pour plusieurs paires de tâches, et obtenu un κ moyen supérieur à 0,4. Trois participants (S10, S03 et S06) ont atteint une moyenne supérieure à 0,7. De plus, un score de classification parfaite a été atteint pour S10 dans les paires ROT-SING et SING-SUB.

Les sept paires les plus performantes étaient toutes formées d'une tâche casse-tête et d'une tâche d'imagerie: ROT ou SUB avec SING, FACE, MI ou NAV. D'autre part, des paires de tâches d'imagerie dynamiques et statiques, soit SING, FACE, NAV et MI, ont montré un κ systématiquement plus faible.

Le κ moyen calculé pour chaque fenêtre d'une seconde est présenté à la Figure 4.1 pour les six meilleures paires de tâches. Les valeurs de κ ont augmenté rapidement après le début du stimulus, et ont atteint un pic près de quatre secondes plus tard. La performance a ensuite diminué progressivement puis est retournée à sa valeur initiale trois secondes après la fin de la tâche.

Les SVM-l_1 ont sélectionné un minimum de zéro et un maximum de 12 traits caractéristiques (médiane de sept). L'évolution du nombre de traits caractéristiques sélectionnés est illustré à la Figure 4.2 pour chaque participant. Au cours des périodes pré- et post-tâche, seulement un ou deux traits caractéristiques ont été sélectionnés. Ce nombre a augmenté rapidement après le début de la tâche, et est resté stable durant ses 15 secondes.

Les traits caractéristiques basés sur les bandes α et β ont montré l'importance la plus forte (voir Tableau 4.5). La plupart de ces traits caractéristiques provenait d'électrodes situées dans les régions pariétales et occipitales. Tel qu'illustré à la Figure 4.3, les canaux dans les régions occipitales et pariétales ont souvent engendré l'importance la plus élevée.

SPIR

La performance de classification maximale des classifieurs entraînés sur des fenêtres d'une seconde de SPIR, pour chaque participant et paire de tâches, est présentée au Tableau 4.6. Le κ moyen le plus élevé a été obtenu à nouveau par S10, avec 13 paires différentes de tâches atteignant une haute performance. D'autre part, S03, qui était classé deuxième avec l'EEG, s'est classé avant-dernier avec la SPIR.

Quinze paires de tâches ont obtenu un κ moyen au-dessus de 0,4. Les combinaisons d'une tâche casse-tête et d'une tâche d'imagerie (ROT-MI, ROT-SING, ROT-FACE, SUB-MI, ROT-WORD) ont de nouveau permis d'atteindre les performances les plus hautes. Les paires combinant deux tâches d'imagerie, telles que SING-MI, NAV-FACE, FACE-MI et SING-FACE, ont systématiquement mené à un κ moyen plus faible.

Tel qu'illustré à la Figure 4.4, les valeurs de κ ont oscillé légèrement au-dessus de 0 au cours de la période de référence et des cinq premières secondes de la tâche. Les valeurs de κ ont ensuite augmenté jusqu'à environ cinq secondes après le début de la tâche, atteignant un plateau près de six secondes plus tard, ce qui correspond au délai attendu introduit par la réponse hémodynamique [9]. La performance a ensuite diminué sur une période d'environ huit secondes, ne revenant pas à son niveau initial dans les cinq secondes après la fin de la tâche.

Les SVM-l_1 ont sélectionné un minimum de zéro et un maximum de 11 traits caractéristiques (médiane de trois). Tel qu'illustré à la Figure 4.5, seul un trait caractéristique a été sélectionné de la période de référence jusqu'à cinq secondes après le début de la tâche, après quoi ce nombre augmente et demeure plus stable jusqu'à la fin de la tâche.

Les sources 13 et 14 (cortex pariétal droit), et les sources 15 et 16 (cortex préfrontaux droit et gauche) ont révélé l'importance la plus élevée (voir le Tableau 4.8). Les traits caractéristiques basés sur HbR ont généralement mené à la plus haute importance, suivis par ceux basés sur HbO, et finalement ceux basés sur HbT. Les canaux situés en région préfrontale se sont révélés particulièrement importants pour les paires utilisant les tâches ROT et SUB, de même que pour les paires avec la tâche WORD en préfrontal gauche (voir la Figure 4.6).

Fusion de l'EEG et de la SPIR

La performance de classification maximale des classifieurs entraînés sur des fenêtres d'une seconde d'EEG et de SPIR, pour chaque participant et paire de tâches, est présentée au Tableau 4.9. Les trois meilleures paires de tâches sont les mêmes qu'en EEG (ROT-FACE, ROT-MI et ROT-SING), tout comme les trois pires (SING-MI, FACE-MI et SING-FACE). Encore une fois, la plupart des paires de tâches les mieux classés étaient une combinaison d'une tâche casse-tête et d'une tâche d'imagerie. De plus, les cinq meilleurs participants (S10, S06, S03, S07 et S05) sont les mêmes qu'en EEG.

Le κ moyen pour des fenêtres d'une seconde est présenté à la Figure 4.7 pour les six meilleures paires de tâches. Deux pics peuvent être observés: d'abord, à trois secondes après le début de la tâche, correspondant au pic en EEG, puis à environ 11 secondes après le début de la tâche, correspondant au pic obtenu avec la SPIR. L'amélioration par rapport à l'EEG semble faible dans les huit premières secondes de la tâche (tel qu'observé au Tableau 4.9, mais est plus marqué dans les cinq dernières secondes de la tâche (voir la Figure 4.5). Puisque cette analyse a pour but d'explorer les façons dont la fusion de l'EEG et de la SPIR peut être utile, le reste de cette section se concentre sur les résultats obtenus dans les cinq dernières secondes de la tâche.

Le Tableau 4.11 rapporte l'augmentation du κ maximal obtenu en ajoutant la SPIR à l'EEG pour la fenêtre allant de 11 à 12 secondes après le début de la tâche, correspondant à la fenêtre optimale en SPIR (voir le Tableau 4.6). L'augmentation moyenne en κ était de 0,17, tandis que l'augmentation maximale était de 0,96, obtenue par S07 pour la paire SUB-MI. À l'opposé, la plus forte baisse était de -0.29, obtenue par S09 avec la paire MOT-NAV. Les gains moyens en performance étaient les plus élevés pour S04 ($\Delta\kappa = 0,27$), S07 ($\Delta\kappa = 0,23$) et S06 ($\Delta\kappa = 0,19$). S10, qui a réalisé la meilleure performance dans les deux conditions unimodales, a bénéficié d'une augmentation moyenne dans le contexte multimodal, mais a subi une diminution de performance pour quelques paires tâches telles que SING-MI et NAV-MI.

Les trois paires de tâches qui ont obtenu les augmentations moyennes les plus élevées ont fait usage de la soustraction mentale: SUB-MI ($\Delta\kappa = 0,41$), SING-SUB ($\Delta\kappa = 0,34$) et SUB-FACE ($\Delta\kappa = 0,30$), trois tâches qui ont également obtenu de bonnes performances en SPIR. Une seule paire (FACE-MI) a subi une diminution de performance moyenne avec l'ajout de la SPIR, tandis que deux paires (NAV-MI et FACE-SING) ont bénéficié d'une augmentation négligeable de 0,01.

Les SVM-l_1 ont sélectionné un minimum de zéro et un maximum de 15 traits caractéristiques (médiane de sept). Tel qu'illustré à la Figure 4.8, l'évolution du nombre de traits caractéristiques sélectionnés est presque identique à celle obtenue en EEG, jusqu'à environ 11 secondes après le début de la tâche, où une légère augmentation du nombre de traits sélectionnées peut être observée.

Finalement, la Figure 4.9 illustre l'importance relative de chaque sous-type de trait caractéristique pour les six meilleures paires de tâches. Les traits caractéristiques basés sur la bande α étaient généralement les plus importants, suivis par la bande β, tandis que les bandes θ, 4 à 30 Hz et 0.1 à 100 Hz étaient moins importants. En SPIR, les traits caractéristiques basés sur HbR ont mené à l'importance la plus élevée (toutefois inférieur à ceux basés sur la bande α), suivis de près par les traits caractéristiques basés sur HbO, alors que ceux basés sur HbT n'ont pas montré beaucoup d'importance.

Effet des fenêtres temporelles sur la performance de classification

L'effet de l'utilisation de différentes longueurs de fenêtres sur les performances de classification est illustré à la Figure 4.10. Des fenêtres temporelles plus longues ont généralement mené à un κ plus élevé en EEG et avec la fusion, à l'exception des fenêtres de 15 s dont la performance était légèrement inférieure à celle

des fenêtres de 10 s. Cependant, la SPIR a vu sa performance décroître de manière constante, voyant sa performance originale coupée de moitié avec des fenêtres de 15 s. De plus, la fusion a mené à des augmentations plus élevées pour les fenêtres plus courtes que pour les plus longues.

0.4.2 Analyse descriptive des données neurophysiologiques

Les données neurophysiologiques ont été analysées afin de mieux comprendre les processus mentaux impliqués dans les tâches étudiées. Notre analyse se concentre sur les configurations spatiales de DLÉ/SLÉ en EEG (voir la Figure 4.11 et de chromophores en SPIR (voir la Figure 4.12). Par souci de concision, les résultats intéressants sont mentionnés en même temps qu'ils sont discutés à la Section 0.5.1.

0.4.3 Analyse du questionnaire d'évaluation subjective de la charge mentale

Comme l'illustre la Figure 4.13, la tâche SUB a induit l'exigence mentale, la pression temporelle, l'effort, et la frustration les plus élevés, a conduit à la plus mauvaise performance perçue, et a été la moins préférée. Les autres tâches casse-tête (WORD et ROT) ont induit une exigence mentale, un effort et une frustration similaires, mais ont été préférées plus souvent. D'autre part, les tâches d'imagerie dynamique (SING, NAV et MI) ont induit une faible exigence mentale et peu de frustration, n'ont pas requis beaucoup d'effort, et ont mené à la perception d'une haute performance. Enfin, la seule tâche d'imagerie passive (FACE), a nécessité une exigence mentale et un effort élevés, a conduit à une performance perçue légèrement inférieure et à un niveau de frustration plus élevé, mais a exercé une faible pression temporelle tout en étant faiblement classée par rapport aux autres tâches.

Corrélation entre traits caractéristiques et évaluations subjectives

Les traits caractéristiques significativement corrélés avec les évaluations subjectives sont présentés aux Tableaux 4.14, 4.15, 4.16, 4.17 et 4.18. La nomenclature utilisée pour nommer les traits caractéristiques est décrite à l'Annexe B.

0.5 Discussion

0.5.1 Tâches et paires de tâches optimales

Le domaine de l'interfaçage humain-ordinateur étant encore récent, plusieurs solutions et concepts restent encore à explorer. Par exemple, la plupart des travaux sur le sujet n'ont étudié que quelques paradigmes tels que les PÉ et PÉVS [14], tandis que des alternatives telles que des tâches mentales non-motrices ont été moins souvent étudiées. Cependant, afin de rendre les ICOs plus flexibles et faciles d'utilisation, de nouveaux paradigmes doivent être identifiés et validés. Dans ce travail, nous avons étudié sept

tâches mentales différentes (rotation mentale, génération de mots, soustraction mentale, chant mental, navigation mentale, imagerie motrice et imagerie de visages), d'un point de vue électrophysiologique et neurohémodynamique, afin de découvrir les plus prometteurs. Dans ce qui suit, les résultats obtenus pour chaque tâche mentale sont discutés et leur potentiel évalué pour une ICO réelle.

Rotation mentale

Les paires de tâches incluant la rotation mentale ont toujours été classées parmi les meilleures paires, avec les trois configurations de modalités (EEG, SPIR et EEG-SPIR). Ces résultats confirment les résultats d'études antérieures [36, 61], dans lesquelles ROT était aussi parmi les meilleures tâches. L'introduction d'un objectif clair dans notre tâche de rotation en fait de plus une tâche casse-tête à l'image de WORD et SUB, dont les performances élevées ont été rapportées précédemment. Cependant, l'augmentation moyenne de la performance apportée par la fusion de modalités s'est avérée parmi les plus basses de toutes les tâches.

Les traits caractéristiques sélectionnés pour les paires utilisant ROT étaient toujours plus importants à l'arrière de la tête en EEG et en préfrontal gauche en SPIR. L'analyse DLÉ/SLÉ illustrée à la Figure 4.11 soutient cette observation: des niveaux élevés de DLÉ ont été observées dans les régions occipitales dans les quatre bandes, ce qui suggère que ces régions ont été recrutées pendant la rotation mentale. Des tendances similaires ont été observées dans les patrons de DLÉ pour la bande $\beta_{inférieur}$ entre 0,5 et 2 secondes après le début de la tâche, de même qu'en région préfrontale, dans [36]. Un patron d'activation des cortex moteurs peut être vu en α et β, ce qui suggère qu'une stratégie impliquant l'imagerie motrice pourrait avoir été utilisée par certains participants pour les aider à imaginer la rotation des solides. Les premiers résultats sont confirmés par une revue de 32 études de neuroimagerie en IRMf et en tomographie par émission de positrons (TEP) qui a conclu que le cortex occipital postérieur était systématiquement activé lors de la rotation mentale [67]. Dans cette revue, l'auteur note également l'activation de régions du cortex préfrontal, ce qui pourrait expliquer la forte baisse en HbO mesuré en région préfrontale.

Bien que classée comme troisième tâche la plus exigeante et frustrante, ROT était souvent parmi les trois tâches préférées des participants. Par conséquent, même si ROT pourrait induire plus de fatigue que d'autres tâches, sa haute performance en EEG et en SPIR, de même que le bon classement donné par les participants, en font un excellent candidat pour une ICO.

Génération de mots

La génération de mots a obtenu une performance de classification supérieure à la moyenne pour les trois configurations de modalités, généralement classé troisième ou quatrième parmi les tâches. Ces résultats sont similaires à ceux observés en EEG par [36]. La fusion s'est avérée particulièrement utile pour WORD, qui a profité de la deuxième augmentation la plus élevée en κ dans un contexte multimodal.

Les traits caractéristiques sélectionnés révèlent une importance particulière dans la région temporale gauche en EEG, et dans la région fronto-temporale gauche en SPIR. En termes de DLÉ/SLÉ, cet effet était surtout perceptible en $\alpha_{inférieur}$ et $\beta_{supérieur}$, où la région temporale gauche démontre des patrons de désynchronisation. Il est également intéressant de noter que les patrons de DLÉ en $\beta_{inférieur}$ dans les régions occipitale et centrale gauche est très similaire à ceux identifiés dans [36]. Les régions temporales ont également été mises en évidence en SPIR, avec une forte augmentation d'HbO du côté gauche et une diminution d'HbR des deux côtés. Encore une fois, ces conclusions sont confirmées par de récentes études en IRMf tentant d'identifier les régions anatomiques nécessaires à la production de la parole [83]. En effet, dans leur revue, Price *et al.* ont constaté que le cortex frontal gauche et la zone de Broca ont été systématiquement activés lors de tâches de récupération de mots. Ces zones du cerveau entrecoupent celles identifiées dans nos résultats.

À l'instar de ROT, WORD a été jugée exigeante et frustrante, mais était toujours parmi les trois tâches préférées par les participants. Toutefois, contrairement à ROT, les participants ont généralement évalué leur performance comme étant faible pour WORD. Ceci pourrait être expliqué par la sélection purement aléatoire de la première lettre des mots devant être générés, ce qui rend la tâche difficile dans certains cas, tels que pour des mots commençant par les lettres Z, X ou Q. Dans l'ensemble, suivant un ajustement de son niveau de difficulté, la génération de mots demeure un bon candidat pour une ICO grâce à sa haute performance en SPIR et en EEG, de même que son évaluation subjective favorable.

Soustraction mentale

La soustraction mentale arrive au deuxième rang en ce qui concerne la performance de classification pour les trois configurations de modalités. SUB est la tâche qui a bénéficié le plus de la fusion: les trois plus fortes hausses de κ ont toutes été enregistrées pour une combinaison de SUB et d'une tâche d'imagerie (MI, SUB et FACE). Ces résultats confirment ceux de [36] en EEG, mais pas de [61] en SPIR. En effet, Hwang *et al.* (2014) ont obtenu des résultats similaires pour la soustraction mentale à ceux d'autres tâches d'imagerie (SING, MI). Cependant, les auteurs ont défini leur tâche comme étant la soustraction successive de deux nombres "simples" (impliquant des nombres à un chiffre) d'un nombre à trois chiffres, ce qui est plus simple que la tâche utilisée dans notre étude. Comme une tâche de multiplication mentale (de deux nombres à deux chiffres) a conduit dans leur étude à la meilleure performance, il est possible qu'une performance similaire ait pu être atteinte avec des soustractions plus difficiles.

Les traits caractéristiques sélectionnés ont révélé l'importance systématique des électrodes PO7 et PO8 en EEG, et des canaux des régions préfrontale droite et temporo-pariétales en SPIR. Ces tendances ont également été observées dans les patrons de DLÉ/SLÉ, avec une forte désynchronisation en α et $\beta_{supérieur}$ dans les régions occipito-pariétale et une partie des régions préfrontales. Des études en IRMf ont identifié l'implication du précuneus, situé dans la partie médiane de la région centro-pariétale, et de régions frontales, comme commune à l'exécution de différentes opérations arithmétiques [84]. Fait intéressant, dans la même étude, Fehr *et al.* ont identifié une activité statistiquement significative des

régions pariétales inférieures bilatérales entre une tâche de soustraction simple et complexe, ce qui pourrait expliquer ces tendances claires d'importance en PO7 et PO8. En ce qui concerne les motifs trouvés en région préfrontale en SPIR, il est possible que la charge accrue sur la mémoire de travail se soit traduite par l'activation du cortex préfrontal dorsolateral [85]. La diminution de HbR dans les deux lobes temporaux étant très similaire à celle observée pour la tâche WORD, il est possible que certains participants aient utilisé une stratégie langagière pour effectuer les soustractions.

En tant que tâche casse-tête, SUB a été classée comme la plus exigeante et frustrante, tout en conduisant à la pire performance perçue et étant la tâche la moins préférée. Avant d'utiliser la soustraction mentale dans une ICO, il serait donc utile d'identifier un niveau de difficulté optimal pour éviter de fatiguer ou de frustrer les utilisateurs. SUB a aussi profité de la fusion EEG-SPIR et est donc recommandé si un tel système est déjà prêt à l'utilisation.

Chant mental

Les paires de tâche utilisant le chant mental ont conduit à une performance inférieure à la moyenne pour les trois configurations de modalités, mais à la troisième plus haute augmentation moyenne de κ en fusion. Des résultats analogues ont été obtenus en EEG [36] et en SPIR [61].

Il est difficile de déterminer si certaines régions du cerveau ont été recrutées systématiquement en se basant uniquement sur l'analyse de l'importance des traits caractéristiques, comme en EEG et en SPIR les cartes topographiques montrent des tendances variables pour les paires utilisant SING. Cette grande variabilité est probablement due à la faible discriminabilité des processus physiologiques sous-jacents au chant mental, en particulier dans les modalités étudiées, qui sont essentiellement limitées aux structures corticales. En effet, si peu d'informations peuvent être décodées en EEG et SPIR, un classifieur mettrait alors l'accent sur les traits caractéristiques fournis par la seconde tâche, et ne choisirait donc pas des traits caractéristiques constants d'une paire à l'autre.

Les études de neuroimagerie portant sur le chant mental ont identifié l'activation de régions fronto-pariétales (cortex moteurs et aire de Broca) [86, 87]. Cela pourrait expliquer les patrons de DLÉ observés dans la bande β dans l'hémisphère gauche, de même que les patrons d'HbR similaires en régions temporale et centrale. Ces études fournissent également une explication de pourquoi le chant mental n'a pas produit de patrons d'activation distincts: d'abord, Gunji *et al.* (2007) ont utilisé la MEG afin de comparer les processus oscillatoires sous-jacents au fredonnement, à la parole, et au chant explicite et implicite (mental), et ont constaté que le chant mental recrutait systématiquement l'aire cérébrale la plus petite [87]. Une petite zone d'activation sera plus difficile à mesurer pour des modalités telles que l'EEG et la SPIR, surtout lorsqu'un faible nombre de capteurs est disponible. Deuxièmement, dans leur étude sur des chanteurs d'opéra, Kleber *et al.* (2007) ont identifié une activation significative des structures profondes du cerveau liées aux émotions telles que l'insula, l'amygdale et l'hippocampe [86]. Il est possible que ces régions aient été activées mais que l'EEG et la SPIR n'aient pu les mesurer comme ils mesurent surtout la surface corticale. La forte SLÉ visible en α et β à la Figure 4.11 supporte cette hypothèse. Cependant,

les patrons de SLÉ étant particulièrement forts dans la région occipitale, il est possible qu'ils soient le résultat d'une synchronisation en α reliée à la faible charge mentale induite (voir la Section 0.5.5).

SING induit la demande mentale, l'effort et la frustration les plus faibles, et a été raisonnablement bien classé par les participants en termes de préférences. Comme les combinaisons de chant mental et d'une tâche casse-tête ont conduit à de hautes performances de classification, il semble raisonnable d'utiliser SING dans une ICO.

Navigation mentale

Les paires de tâche utilisant la navigation mentale ont mené à une performance inférieure à la moyenne pour les trois configurations de modalités, semblable en termes de performance au chant mental, et ont mené à la plus faible augmentation de κ en fusion. Des résultats similaires ont été obtenus en EEG seul [36].

Les patrons de DLÉ/SLÉ observés pour la navigation mentale sont caractérisés par l'activation de l'hémisphère gauche, en particulier dans les bandes $\alpha_{inférieur}$ et $\beta_{inférieur}$. Ces résultats confirment ceux de [36], et sont de plus soutenus par les résultats d'une étude IRMf sur 16 participants exécutant une tâche de navigation mentale [88] . En effet, Ino *et al.* (2002) ont rapporté la présence d'activations statistiquement significatives dans la région prémotrice gauche, le gyrus angulaire gauche (lobe pariétal) et les structures cérébrales plus profondes telles que les zones rétrospléniales bilatérales, certaines parties de l'hippocampe et le cervelet [88]. Toutefois, ces patrons ne sont pas identifiables dans notre analyse de l'importance des traits caractéristiques, où différentes régions du cerveau semblent utiles pour chaque paire autant en EEG qu'en SPIR, comme ce fut le cas pour SING. Toutefois, Ino *et al.* ont utilisé une tâche de navigation plus complexe demandant en plus de compter le nombre de virages effectués. Comme notre étude et celle de Friedrich *et al.* ont toutes deux mené à une mauvaise performance de cette tâche, il est possible que l'augmentation du niveau de difficulté puisse contribuer à produire des activations plus fortes et plus claires.

La navigation mentale a été classée comme la troisième tâche la moins exigeante et frustrante, et a généré une haute performance perçue. Classée quatrième tâche préférée en moyenne, NAV, tout comme SING, peut être considérée pour une ICO, surtout parce qu'elle a conduit à de bonnes performances lorsque jumelée avec des tâches casse-tête. Cependant, sa faible augmentation en κ dans le cas multimodal la rend moins utile à utiliser avec une approche SPIR-EEG.

Imagerie motrice

L'imagerie motrice est sans aucun doute la tâche mentale la plus souvent utilisée dans la littérature [14], fournissant une multitude d'articles auxquels comparer nos résultats. Alors que MI a été la deuxième pire tâche en SPIR, elle a mené aux troisième et quatrième meilleures performances en EEG et en fusion, respectivement. MI a également été la deuxième tâche bénéficiant le moins de la fusion. Bien qu'utilisant

une méthodologie différente, Fazli *et al.* (2012) ont étudié un système hybride EEG-SPIR similaire au nôtre, mais seulement évalué l'utilisation d'un paradigme basé sur l'imagerie ou l'exécution motrice de la main gauche ou droite [17]. Les auteurs ont rapporté une augmentation moyenne de la précision de classification de 5% avec la fusion de l'EEG et de la SPIR dans 90 % des sujets, difficile toutefois à comparer à notre étude en raison de différences méthodologiques. Cependant, nous avons également noté une nette amélioration de performance chez tous les participants ($\Delta\kappa = 0,13$), même si cette amélioration était parmi les plus basses de nos sept tâches.

Bien que les patrons d'activation induits par MI aient probablement été masqués par ceux de tâches casse-tête dans la plupart des paires (comme ce fut le cas pour SING et NAV), l'analyse de l'importance des traits caractéristiques montre la prédominance des cortex moteurs gauche et droit pour les paires NAV-MI, SING-MI et FACE-MI en EEG. Ce phénomène semble se produire uniquement pour la paire NAV-MI en SPIR, ce qui pourrait plutôt révéler un patron propre à NAV, comme la paire NAV-FACE montre des traits caractéristiques importants similaires. Les patrons d'activation DLÉ/SLÉ dans les trois premiers groupes ont en outre permis d'identifier l'activation du cortex moteur, et la prédominance de l'hémisphère gauche. De nombreuses études sur les ICO ont fait état de patrons similaires [89, 90], et le patron obtenu en $\beta_{inférieur}$ est très similaire à celui rapporté dans [36]. Cela est en outre soutenu par de fortes augmentations en HbO et HbT pour l'hémisphère gauche. Comme pour SING, une forte SLÉ dans les bandes α, principalement dans le lobe occipital, suggère que la charge mentale plus faible induite par cette tâche pourrait produire des artéfacts liés à la somnolence (voir la Section 0.5.5) .

L'imagerie motrice a induit un exigence mentale, effort et une frustration parmi les plus bas, a généré la plus haute performance perçue, et a été classée comme la tâche la plus agréable en moyenne. Les combinaisons de MI avec une tâche casse-tête a de plus conduit à de hautes performances de classification. MI est donc une tâche utile qui mérite sa place prédominante dans la littérature.

Imagerie de visages

L'imagerie de visage était la pire tâche dans toutes les configurations de modalité. La fusion a permis d'améliorer la performance de classification, de manière semblable à la plupart des autres tâches. Ces résultats corroborent ceux de Friedrich *et al.* en EEG [36].

La principale structure cérébrale recrutée lors de l'exposition visuelle à un visage est le gyrus fusiforme, situé dans la partie postérieure du cortex [91]. Cette structure étant loin des couches corticales pouvant être mesurées en EEG et en SPIR, il est logique que la classification de l'imagerie de visage soit une tâche difficile. Cela peut aussi expliquer les patrons de traits caractéristiques disparates observés à la fois en EEG et en SPIR.

Le patron de DLÉ/SLÉ observé dans la bande $\beta_{inférieur}$ était similaire à celui rapporté dans [36]: une désynchronisation des lobes frontaux et centraux, et une synchronisation des lobes temporaux. En outre, les bandes α montrent une forte SLÉ en région occipitale, qui pourrait à nouveau être expliquée par la somnolence induite par une faible charge mentale.

Bien qu'induisant une faible pression temporelle, la tâche d'imagerie de visage était presque à égalité avec les tâches casse-tête pour ce qui est de l'exigence mentale et de l'effort, suggérant que les participants ont trouvé difficile d'imaginer le visage de leur ami. Étant de plus parmi les trois tâches les moins biens classées en termes de préférence, l'imagerie de visage n'est pas un bon candidat pour une ICO, car elle souffre de faibles performances unimodale et multimodale, et n'a généralement pas été appréciée par les participants.

Paires de tâches optimales

En se basant sur ces résultats, nous concluons que, sur les 21 paires étudiées, une combinaison formée d'une tâche casse-tête (ROT, SUB ou WORD) et d'une tâche d'imagerie (MI, FACE, NAV, SING) est la plus susceptible d'atteindre une bonne performance. Cela confirme les résultats de [36] et [61]. De plus, dans le cas où un système EEG-SPIR est disponible, nos résultats montrent que les paires utilisant SUB et WORD pourraient voir leur performance grimper le plus lorsque la fusion de modalité est utilisée.

0.5.2 Évaluation de la performance dans un contexte réel d'utilisation

Quelques points méthodologiques devront être abordés différemment pour une implantation en temps réel du paradigme d'ICOh décrit plus haut. En premier lieu, nous avons choisi de mesurer l'augmentation des performances pour une fenêtre temporelle allant de 11 à 12 secondes après le début de la tâche afin de mettre en évidence la valeur ajoutée de la fusion. Cette approche devra être adaptée lors de pour une application en temps réel. Par exemple, plusieurs études en SPIR ont extrait des traits caractéristiques sur des fenêtres de différentes longueurs et les ont combinés pour ne former qu'un seul jeu de données [56–58, 61, 62]. Bien que cette approche puisse nuire au classifieur en augmentant la dimensionalité de l'entrée, elle fournit également plus d'informations et peut donc conduire à une meilleure performance.

Un autre point critique mis en évidence dans la plupart des études mentionnées ci-haut est le délai introduit par la réponse hémodynamique en SPIR. En effet, dans notre étude, nous avons constaté qu'environ 11 secondes étaient nécessaires pour que la SPIR atteigne sa performance maximale, rendant du coup l'utilisation d'une ICO unimodale en temps réel suboptimale. D'un autre côté, certains participants ont vu leur performance en EEG décroître à travers les 15 s de réalisation d'une tâche, probablement dû à l'ennui ou la fatigue, nuisant ainsi à la qualité de leur imagerie mentale. En combinant la SPIR avec l'EEG, il est possible d'atteindre une performance satisfaisante dans les premières secondes après le début de la tâche, tout en bénéficiant de performances accrues dans les secondes suivantes.

Malgré ces limitations, certaines applications pourraient bénéficier des particularités de ce paradigme d'ICOh. Par exemple, une ICOh requérant le maintien d'un état mental pendant quelques secondes permettrait d'assurer un certain niveau de certitude avant qu'une décision finale ne soit prise et émise par l'interface. Notre système serait ainsi utile pour reconnaître une tâche mentale spécifique sur une plus longue période de temps, tout en offrant une performance de classification accrue. En outre, un

utilisateur souffrant d'un contrôle médiocre en EEG pourrait tout de même être en mesure d'atteindre une performance élevée, ce qui s'est révélé être le cas pour le participant S04 de notre étude.

0.5.3 Effet de la fenêtre temporelle

L'effet de la sélection de la fenêtre temporelle sur la performance de classification, tel qu'illustré à la Figure 4.10, révèle une tendance croissante de la performance en EEG et en fusion, mais pas en SPIR. En effet, des fenêtres temporelles plus longues ont produit des performances de classification plus élevés en EEG et en fusion, ce qui est attendu sachant que des fenêtres plus longues vont généralement incorporer plus d'information sur les processus temporels qui se produisent au cours de la tâche. Cependant, des fenêtres plus courtes en SPIR ont étonnamment conduit à de meilleures performances que celles plus longues. Ceci pourrait être expliqué par le fait que les niveaux moyens de chromophore sont moins informatifs lorsque moyennés sur une période incluant les premières secondes de la tâche, où peu d'informations utiles sont disponibles en raison du délai de la réponse hémodynamique.

Parce que le temps de réponse d'une ICO et sa performance de classification multimodale augmentent tous deux avec des fenêtres plus longues, le choix d'une longueur de fenêtre temporelle revient ainsi à un compromis entre la réactivité et la précision du système. Dans notre cas, nous avons vu que les fenêtres d'une seconde peuvent conduire à une bonne performance de classification dans la plupart des sujets, tout en bénéficiant de la plus forte augmentation en κ pour la fusion. Dans les applications où un délai est acceptable, il pourrait néanmoins être plus utile d'utiliser des fenêtres plus longues.

0.5.4 Analyse de la corrélation

L'analyse de la corrélation entre les caractéristiques neurophysiologiques extraites et les évaluations recueillies à l'aide du questionnaire NASA TLX devrait permettre de dévoiler des marqueurs potentiels de l'expérience subjective, fournissant des informations utiles pour la conception d'ICO plus intelligentes.

Tout d'abord, notre expérience n'a pas été conçue pour fournir une évaluation juste des dimensions subjectives. En effet, bien que les évaluations aient été collectées trois fois pour chaque participant et tâche mentale, le niveau de difficulté n'a pas été varié: cela signifie qu'il est possible, si les évaluations demeurent essentiellement les mêmes pour chaque tâche mentale, que ces résultats proviennent de différences inhérentes entre les tâches et non pas de la charge qu'elles induisent chez les participants. Par conséquent, afin de valider ces résultats, il serait nécessaire de mener des études supplémentaires où la difficulté de la tâche est contrôlée. Cependant, en comparant nos conclusions avec les résultats de l'analyse de chaque tâche discutés ci-haut, il est possible de poser une hypothèse sur la probabilité que ces résultats aient été causées par les différences entre les tâches ou par d'autres processus physiologiques.

L'exigence mentale et l'effort, étant corrélés avec un ρ_{Spear} de 0,79, ont tous deux montré une corrélation négative avec les niveaux de HbR dans la région centro-frontale gauche. L'effet inverse a toutefois été observé à la Figure 4.12), où les tâches induisant une haute exigence mentale (ROT,

WORD, SUB et FACE) ont en fait conduit à des niveaux plus élevés de HbR. Comme indiqué plus haut, le recrutement de cette région peut être expliqué par l'augmentation de la charge de mémoire de travail causée par ces tâches, engendrant une activation du cortex préfrontal dorsolateral [85].

La pression temporelle était négativement corrélée avec la modulation d'amplitude en α dans la région pariéto-occipitale, et avec de nombres traits caractéristiques basés sur les niveaux d'HbO en région préfrontale. Ce second patron d'activation correspond à la SLÉ observée en pariéto-occipital dans les bandes α et β pour les tâches à la pression temporelle faible (SING, MI et FACE). Quant à la SPIR, la diminution d'HbO en région préfrontale avec ROT, SUB et WORD (de même que l'augmentation d'HbO dans les autres tâches, en particulier MI) pourraient expliquer cet effet. Par conséquent, il semble que les corrélations significatives observées pour la demande temporelle puissent être expliquées par les différences entre les paires de tâches.

Quant à la performance perçue, plusieurs traits caractéristiques de modulation d'amplitude en α et β ont montré une corrélation positive en région centro-pariétale. Les tâches induisant de plus faibles performances (WORD, SUB et FACE), cependant, ont montré des tendances très différentes en DLÉ/SLÉ, comme ce fut le cas pour les tâches induisant de hautes performances (ROT, SING, NAV et MI). L'effet observé par l'analyse des corrélations pourrait donc être spécifique à la modulation d'amplitude en α et β, hypothèse soutenue par le fait qu'aucun des traits caractéristiques classiques comme les bandes de puissance n'aient été sélectionnés. En outre, une étude en IRMf sur neuf participants a montré une augmentation du signal BOLD en fronto-pariétal dans les réalisations de tâches où la bonne réponse à une tâche de mémoire visuelle était donnée [92], une région qui entrecoupe celle sélectionnée dans notre analyse. Par conséquent, l'effet observé pourrait bien être liée à la performance perçue, bien que d'autres expériences doivent être faites pour valider cette hypothèse.

Enfin, deux traits caractéristiques étaient significativement corrélés avec le niveau de frustration, en régions centrale et centro-pariétale. L'effet de la frustration a été évaluée en IRMf dans [93], révélant le recrutement de l'amygdale, d'une partie du mésencéphale, de certaines parties profondes du cortex (insula) et du cortex préfrontal. Le manque de correspondance entre nos résultats et la littérature indique que d'autres expériences doivent être menées pour évaluer correctement l'effet de la frustration sur les signaux physiologiques en EEG et SPIR.

Dans l'ensemble, notre analyse exploratoire des corrélations a montré des résultats prometteurs qui pourraient être utilisés pour compléter les fonctionnalités d'une ICO. Néanmoins, de plus amples travaux avec une méthodologie conçue à cet effet seront nécessaires pour confirmer ces conclusions.

0.5.5 Limitations

À la lumière des résultats décrits ci-dessus, certains éléments nécessitant des ajustements méthodologiques ou une étude plus poussée valent la peine d'être mis en évidence.

Tout d'abord, en ce qui concerne le paradigme expérimental, il serait utile de tenir compte de la variabilité introduite dans la mesure de signaux neurophysiologiques par les casques d'EEG-SPIR utilisés. Pour ce faire, il serait possible de numériser les emplacements de chaque sonde par rapport à certains repères anatomiques sur la tête des participants, et d'utiliser cette information pour reproduire un modèle plus fidèle des données recueillies. Deuxièmement, l'utilisation de nouvelles technologies de casques et de capteurs pourrait rendre le processus de collecte de données plus efficace, en diminuant par example le temps d'installation. Enfin, un biais potentiel de notre étude provient de la somnolence induite par certaines tâches, tel qu'indiqué par les analyses topographiques soulignant l'importance de la bande α dans les régions occipitale et pariétale postérieure. Cela pourrait être causé par la somnolence induite par les tâches qui ne nécessitent pas beaucoup d'effort mental, menant certains participants à fermer momentanément leurs yeux sous l'effet de la fatigue ou de l'ennui. En effet, il a été démontré que la forte synchronisation en α est générée par les circuits neuronaux situés dans les régions occipitale et pariétale postérieure du cerveau lorsque l'on ferme les yeux [94]. Le protocole expérimental devrait donc être amendé pour éviter de telles périodes de somnolence. Par exemple, il serait possible de donner plus de temps aux participants entre deux sous-sessions pour qu'ils se reposent, les aidant à se concentrer de nouveau pendant la suite de l'expérience.

Ensuite, d'autres considérations concernant la procédure de classification devraient être prises en compte. Tout d'abord, une seule valeur de l'hyperparamètre C a été utilisée pour tous les modèles. Ce choix a été fait de sorte à ce que l'analyse puisse se concentrer sur l'évolution de κ à travers le temps, sur les traits caractéristiques sélectionnées, de même que sur l'impact de l'ajout de la SPIR à l'EEG, tout en évitant une division supplémentaire des données nécessaire pour choisir une valeur d'hyperparamètre optimale et ensuite la tester sur une autre partition des données. Cependant, ce choix signifie que les valeurs de κ ne sont pas optimales. En effet, chaque combinaison de participant, paire de tâches et de fenêtre temporelle représente un problème de classification distinct, et un hyperparamètre différent pour chaque classifieur entraîné pourrait mener à une meilleure performance. Une solution à ce problème est d'utiliser une validation croisée imbriquée pour choisir entre plusieurs valeurs d'hyperparamètres ou algorithmes (par exemple, analyse discriminante linéaire, régression logistique, etc.). Deuxièmement, trois procédures ont été utilisées pour minimiser le problème de la malédiction de la dimensionalité: 1) se concentrer sur les traits caractéristiques standards tels que les bandes de puissance et l'amplitude des variations de chromophores, 2) utiliser les SVM, qui sont connues pour leur robustesse en haute dimensionalité, et 3) utiliser une méthode de régularisation l_1 qui ne sélectionne que les caractéristiques les plus pertinentes tout en évitant un partitionnement supplémentaire des données. En dépit de ces procédures, la dimensionalité de l'ensemble d'apprentissage pour la fusion de modalités était d'environ 777 traits caractéristiques pour 86 exemples. D'autres options auraient pu être considérées: certaines méthodes de réduction de la dimensionalité telles que l'analyse en composantes principales (ACP), la collecte d'autres points de données (requérant toutefois un important investissement de temps et d'argent) ou la classification inter-sujet, permettant de combiner des exemples de tous les neuf participants ensemble, multipliant ainsi la quantité de données par neuf.

0.6 Conclusion

Dans ce travail, nous avons étudié l'utilisation de deux techniques de neuroimagerie fonctionnelle non invasives, l'EEG et la SPIR, pour la classification binaire de sept différentes tâches mentales. Les paires de tâches optimales en EEG, en SPIR et en fusion EEG-SPIR, de même que l'impact d'une approche multimodale sur la performance de classification, ont été évalués chez neuf participants.

Les paires formées d'une tâche casse-tête, c'est-à-dire qui requiert la résolution d'un problème (ROT, SUB, WORD), et une tâche d'imagerie (MI, FACE, SING) a systématiquement mené à la meilleure performance de classification en EEG et en SPIR, ainsi que pour la fusion de ces deux modalités. Alors que la performance unimodale obtenue est du même ordre que celle d'études précédentes, l'approche multimodale a conduit à une augmentation moyenne de 0,03 du κ de Cohen maximal lors de l'utilisation des traits caractéristiques extraits de fenêtres d'une seconde (équivalent à une augmentation de la précision de 1,5% pour des classes équilibrées). De même, une augmentation de 0,17 du κ (augmentation de la précision de 8,5%) a été obtenue pour les fenêtres extraites dans les cinq dernières secondes de l'exécution de tâches mentales.

L'analyse des classifieurs entraînés a dévoilé des patrons d'activation spatiale intéressants et l'importance de différents sous-types de traits caractéristiques dans la fusion de modalités. En particulier, les bandes α et β en EEG se sont avérées les plus utiles, suivies par les variations d'HbR et d'HbO en SPIR. Les traits caractéristiques les plus importants en EEG provenaient des régions occipitales et pariétales, tandis que pour la SPIR les régions préfrontale et frontale se sont avérées les plus informatives. L'évaluation de la charge générée par les tâches recueillie auprès des participants à chaque session d'enregistrement a indiqué une préférence pour l'imagerie motrice, la rotation mentale et la génération de mots, et a confirmé la charge mentale élevée induite par les tâches casse-tête. Enfin, certains traits caractéristiques d'EEG et de SPIR qui pourraient être utilisés pour prédire l'expérience subjective d'un utilisateur d'ICO ont été identifiés grâce à une analyse des corrélations, même si d'autres recherches seront nécessaires pour valider ces résultats.

Afin de permettre la conception d'ICOh EEG-SPIR fonctionnelles en temps réel, des étapes supplémentaires peuvent être nécessaires. Tout d'abord, la performance de classification peut être améliorée en incluant des traits caractéristiques supplémentaires, en utilisant des informations complémentaires provenant de signaux physiologiques périphériques (comme la fréquence cardiaque, la respiration, la réponse électrodermale, etc.), et la sélection de modèles de classification optimaux par l'optimisation des hyperparamètres et des algorithmes utilisés. L'effet de la combinaison de différentes longueurs de fenêtres temporelles devrait aussi être pris en compte pour optimiser le compromis entre temps de réponse et précision. Deuxièmement, la classification inter-sujet, plutôt qu'intra-sujet, devrait être explorée afin d'établir la faisabilité d'une classification universelle, qui possède beaucoup d'avantages pratiques. Troisièmement, la faisabilité du protocole multimodal doit être évaluée sur une population plus diversifiée, comprenant des personnes handicapées qui pourraient bénéficier de cette technologie, afin de s'assurer que nos résultats sont transférables et généralisables. L'étape-clé, cependant, demeure la mise en œuvre

d'ICO fonctionnelles en temps réel basées sur les paires de tâches optimales identifiées dans ce travail. Les restrictions causées par une implantation en temps réel (temps d'installation, temps de calcul et latence, bruit ambiant, etc.) devront être surmontées pour produire une interface cerveau-ordinateur utile en pratique.

Contents

List of Figures

List of Tables

Liste des figures

Liste des tableaux

List of Abbreviations

BCI	Brain-Computer Interface
CCA	Correlation Components Analysis
CNS	Central Nervous System
CSP	Common Spatial Patterns
ECG	Electrocardiography
ECoG	Electrocorticography
EEG	Electroencephalography
EMG	Electromyography
EOG	Electrooculography
ERD	Event-Related Desynchronization
ERP	Event-Related Potential
ErrP	Error Potential
ERS	Event-Related Synchronization
FACE	Face imagery
FES	Functional Electrical Stimulation
fMRI	Functional Magnetic Resonance Imaging
fNIRS	Functional Near-Infrared Spectroscopy
fTCD	Functional Transcranial Doppler
GFS	Global Phase Synchrony
hBCI	Hybrid Brain-Computer Interface
HbO	Oxy-hemoglobin
HbR	Deoxy-hemoglobin
HbT	Total hemoglobin

HRF Hemodynamic Response Function

ITR Information Transfer Rate

κ Cohen's kappa inter-rater agreement coefficient

LDA Linear Discriminant Analysis

l_1-**SVM** Support Vector Machine with l_1 regularization

MBLL Modified Beer-Lambert Law

ME Motor execution

MEG Magnetoencephalography

MI Motor imagery

NAV Mental navigation

OD Optical Density

PRB Practical Bit Rate

PCA Principal Components Analysis

PET Positron Emission Tomography

PPG Photoplethysmography

ρ_{spear} Spearman's rank correlation

ROT Mental rotation

SCP Slow Cortical Potential

SCR Skin Conductance Response

SING Mental singing

SSEP Steady-State Evoked Potential

SSAEP Steady-State Auditory Evoked Potentials

SSSEP Steady-State Somatosensory Evoked Potentials

SSVEP Steady-State Visually Evoked Potentials

SUB Mental subtraction

SVM Support Vector Machine

WORD Word generation

Chapter 1

Introduction

1.1 Motivation

A Brain-Computer Interface (BCI) is a communication system between a brain and a computer that bypasses the normal brain output pathways such as muscle movement [1]. Such systems can be useful to replace, restore, enhance, supplement, or improve the natural output of the central nervous system [2], and have found applications in clinical as well as non-clinical contexts such as entertainment and education [3]. In particular, a significant body of research is focused on applying these systems to help individuals with disabilities, such as people suffering from paralysis induced by a degenerative disease (e.g., amyotrophic lateral sclerosis) or caused by an accident [4]. Indeed, these individuals often cannot communicate or interact with their environment due to their condition, and a BCI becomes the only way they can do so, potentially leading to a drastically improved quality of life.

Although most of today's BCI designs use Electroencephalography (EEG) alone to recognize user intent (see Section 2.2.1), other modalities offer different information about the underlying brain activity and can therefore complement the information obtained with EEG alone. Indeed, Magnetoencephalography (MEG), Functional Near-Infrared Spectroscopy (fNIRS), Functional Magnetic Resonance Imaging (fMRI), and many others [5–11] have already been used for that purpose. The multimodal BCI approach consists in using more than one modality at a time, including at least one brain modality, but possibly including non-neurophysiological modalities as well, to improve on the performance and usability of a unimodal system [12, 13]. This approach is a subtype of the more general Hybrid Brain-Computer Interface (hBCI) approach, that combines different mental activity patterns and/or modalities to achieve that effect.

To allow volitional control of a device, BCIs typically rely on the recognition of one or multiple distinguishable brain activity patterns that can be voluntarily produced by the user. Through their extensive use in the literature, specific brain activity patterns such as Event-Related Desynchronization (ERD) and Event-Related Synchronization (ERS), the P300 Event-Related Potential (ERP) and the

Steady-State Visually Evoked Potentials (SSVEP), have all been shown to be distinguishable when used in BCI designs; however, they may not be optimal for all BCI users. First of all, inter-subject variability, a phenomenon that describes how neurophysiological signals can differ significantly from an individual to another, implies that certain patterns will be more discriminant than others - and thus better suited - for some users while they lead to poor performance in others [14]. Finding the optimal set of brain activity patterns for a user can thus significantly improve the performance and usability of a BCI. Secondly, users who have suffered a brain injury may lose normal functioning of regions of their brain associated with the above mentioned patterns. For these target users of BCI systems, it is necessary to identify a variety of different brain activity patterns that recruit other regions of the brain.

The present work aims at identifying - and by doing so, at validating previous research on - mental tasks that can be effectively used to control an EEG, NIRS or EEG-NIRS BCI, while investigating the usefulness of these modalities toward creating highly functional BCIs.

1.2 Research Objectives and Contributions

To address the potential value of different mental task for the control of a BCI, we asked the following questions:

1. What mental tasks can be useful to control a BCI with a minimal usable accuracy of 70% as recommended in [15]?
2. In which context could a multimodal EEG-NIRS BCI provide significantly better performance than a unimodal EEG- or NIRS-only BCI? Moreover, could these modalities provide supplementary information that makes for more user-friendly BCIs?

Previous research found EEG and NIRS to be useful in classifying various mental tasks (see Section 2.3). Most of these studies were able to identify pairs of mental tasks that reached a usable level of 70% accuracy, however, few of them looked at more than three mental tasks. Moreover, these studies often neglected to analyze the behaviour of the classification pipeline used to distinguish between mental tasks, overlooking valuable insights about the relative importance of each type of feature and their spatial location [16]. In addition, only three studies [17–19] (described in Section 2.3) investigated the fusion of EEG and NIRS information for classifying mental tasks. Together, these studies only looked at three types of mental tasks (motor imagery, mental subtraction and mental counting), thus leaving open the question of how fusion can benefit to other types of mental tasks. To fill these gaps, the objectives of this thesis are to:

1. Compare and analyze the classification accuracy and overall performance (including subjective appreciation) of different pairs of mental tasks using EEG alone, NIRS alone, or a multimodal NIRS-EEG BCI.

2. Evaluate the explanatory power of different neuroelectrophysiological and neurohemodynamic features in regard to the subjective experience of a BCI.

Part of the present work has been submitted or is currently being prepared for submission to peer-reviewed publications. The introductory content presented in Section 2.2 is taken from a comprehensive review of the hBCI literature, which has been accepted pending minor revisions to the Journal of Brain-Computer Interfaces [14]. Furthermore, results from the EEG-NIRS fusion experiment presented in Chapter 4 make up the core material of a paper in preparation and to be submitted to the IEEE Transactions on Neural Systems and Rehabilitation Engineering [95].

1.3 Thesis Outline

The present chapter has introduced the challenge of improving BCI performance through the identification of new mental tasks and the use of additional modalities, and highlighted the main research questions and objectives of this work. Chapter 2 covers the rationale behind the research project, including an overview of the field of brain-computer interfacing, relevant results from a comprehensive review of the literature on hBCIs, and suggested improvements to the state-of-the-art. Chapter 3 describes the chosen methodology, from the neurophysiological signals collection to the processing and analysis pipelines. This is followed by the presentation of the results in Chapter 4, and their discussion in light of the current state-of-the-art in Chapter 5. Finally, a summary of this work is presented and future research directions are outlined in the conclusion (Chapter 6).

Chapter 2

Brain-Computer Interfaces: Recent Advances and Open Challenges

This chapter introduces the terminology and the main concepts behind brain-computer interfacing. First, background material and the main results from a systematic review of the hybrid brain-computer interface literature are included in Sections 2.1 and 2.2 to lay out the state-of-the-art as well as current challenges[1]. Second, the use of mental tasks in unimodal and multimodal BCIs is introduced through a brief survey of the literature in Section 2.3. Finally, the proposed improvements brought by this work are introduced in Section 2.4.

2.1 Structure, types and definitions

A BCI is a communication system that bypasses the normal brain output pathways [1]. While this type of system was initially intended for individuals with severe motor impairments [20], applications for healthy users were also proposed, for example in videogames [21, 22] and affective monitoring [23]. BCIs rely on imaging modalities that can record brain activity, such as EEG, MEG, fNIRS, fMRI and Electrocorticography (ECoG) [5–10, 24]. EEG BCIs, in particular, have been the subject of most of the work in this field thanks to their well-studied features, affordability, portability and high temporal resolution [1, 6].

Typically, a BCI works by recognizing one of a few distinguishable brain activity patterns, or *neural response patterns*, that can be elicited by the user internally (consciously) or unconsciously, by using external stimuli [24]. In the case of EEG BCIs, the most commonly used neural response patterns are ERPs, Steady-State Evoked Potentials (SSEPs) or Slow Cortical Potentials (SCPs) [96]. Other types of

[1] This work, entitled "Recent advances and open challenges in hybrid brain-computer interfacing: A technological review" [14] has been submitted to the Brain-Computer Interface journal and has been accepted pending minor revisions.

BCIs work similarly by requiring the elicitation of specific mental activity patterns that can be recognized via the modality of interest. Recently, a new approach called hBCI has emerged, which seeks to combine two or more BCIs, or a BCI with another input modality such as biosignals or traditional human-computer interface devices [13, 25]. The main goal in designing hBCIs is to use a particular modality to overcome the limitations of another. Hybrid BCIs are promising due to their enhanced flexibility (allowing the interface to adapt to the BCI proficiency of different subjects), increased robustness (the combination of inputs can yield better recognition of user intent) and higher number of possible commands. For the sake of completeness, in the sections to follow, structures of BCIs and hBCIs are presented, along with their subtypes and formal definitions.

2.1.1 Measuring neurophysiological activity

A brain-computer interface consists of multiple modules forming a closed loop between the user and the computer [24, 26], as depicted by Figure 2.1. First, a brain imaging modality is used to collect information on the user's mental activity. EEG, NIRS and Functional Transcranial Doppler (fTCD) use probes that are placed on the user's scalp to measure electrical potentials arising from neuronal activity in the former case, and to measure the resulting cerebral hemodynamic changes in the two latter cases. MEG measures magnetic potentials through sensors statically located around the user's head and requires a large device restricted to use in a controlled environment [97]. Similarly, fMRI requires an expensive and static setup in a controlled environment, to measure brain activity through changes in cerebral blood flow [98]. Unlike these non-invasive techniques, ECoG uses electrodes positioned directly on the subject's cortex, which requires surgical procedures [7]. In hybrid BCIs, supplementary modalities may be used in addition to the aforementioned ones: for example, Electromyography (EMG), Electrooculography (EOG), Electrocardiography (ECG), Photoplethysmography (PPG), Skin Conductance Response (SCR) sensors, skin temperature sensors, blood pressure sensors, respiration rate sensors, eye trackers and traditional human-computer interface sensors (keyboard, game controller, analog sensor, etc.) [13]. EMG, EOG and ECG collect biopotentials arising from different parts of the body, namely, from muscles, eye movements, and the heart, respectively. A PPG sensor, in turn, optically monitors the local changes in microvascular blood volume and is commonly used for measuring heart rate [99]. Lastly, SCR measures the electrical conductivity of the skin, which is known to vary with different affective states and with autonomic activity [100]. We now describe in more detail the two modalities most often used in BCIs, and that are the subject of this work: EEG and NIRS.

Electroencephalography

The brain is composed of millions of billions (10^{11}) of cells called neurons [27]. These cells are characterized by a unique ability to convey and transmit information via electrochemical activity.

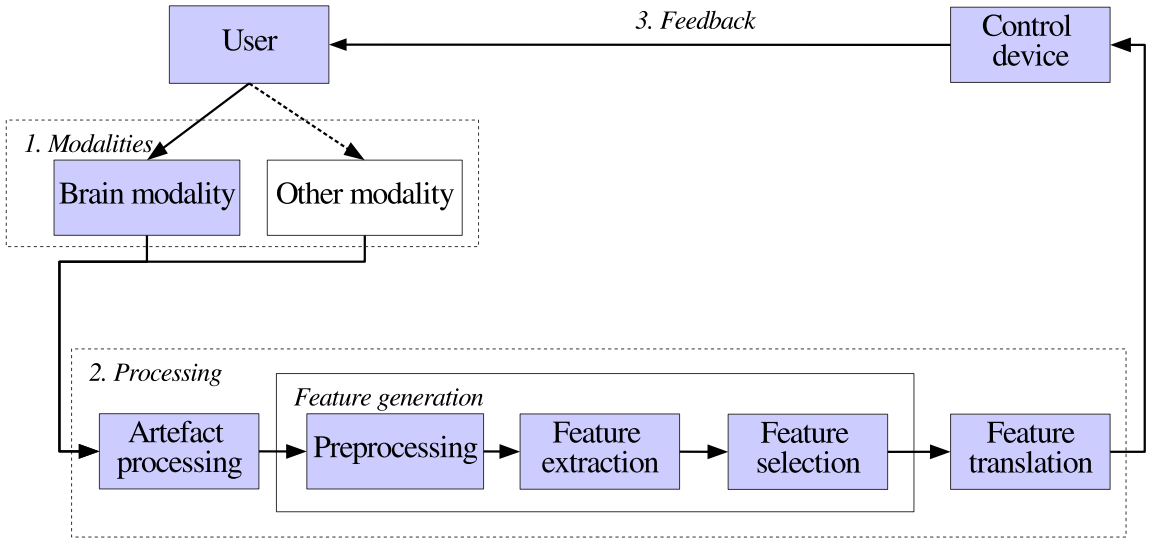

Figure 2.1: Structure of a standard hBCI. This functional model is inspired by [26] and [24].

Each neuron receives a vast amount of upstream information from other neurons or from different sensory cells [101]. Between neurons, this information is generally transmitted via messenger molecules called neurotransmitters that are released by one neuron, and detected by downstream neurons. These neurotransmitters can either inhibit, excitate or modulate the activity of the neurons that detect them. At any moment, a neuron receives and integrates information from many different neurons, firing only if these inputs exceed the neuron's threshold potential. This firing, called action potential, is characterized by the propagation of an electrical current along the membrane of the neuron's main output, the axon. Once the current reaches the axon terminal, neurotransmitters are released onto the downstream neurons, and so on. This electrochemical event allows rapid transmission of information along the axon to other downstream neurons or cells (such as muscle cells). From resting to peak firing, the electrical potential across a single neuron's membrane varies between -60 and 50 mV and can thus be measured using micro-electrodes implanted inside or near the neuron.

In the cortex, the highly circumvoluted outer layer of the brain that manages higher order functions, neuron populations are highly structured. Six layers of neurons, characterized by their size and shape, can be distinguished [102]. Because of this ordered architecture where the axons are locally oriented in the same direction, perpendicular to the scalp, the electrical potentials produced by a population of neurons add up instead of cancelling out. If this firing occurs in synchrony, an electrical potential of approximately 100 μV can be measured on the scalp [27]. This means that a measure taken at some location on the scalp is in fact the summation of a great number of different electrical fields, originating from different parts of the brain and attenuated in the heterogeneous medium of brain tissue, meninges, skull and skin. EEG thus offers a poor spatial resolution: an EEG signal does not contain information

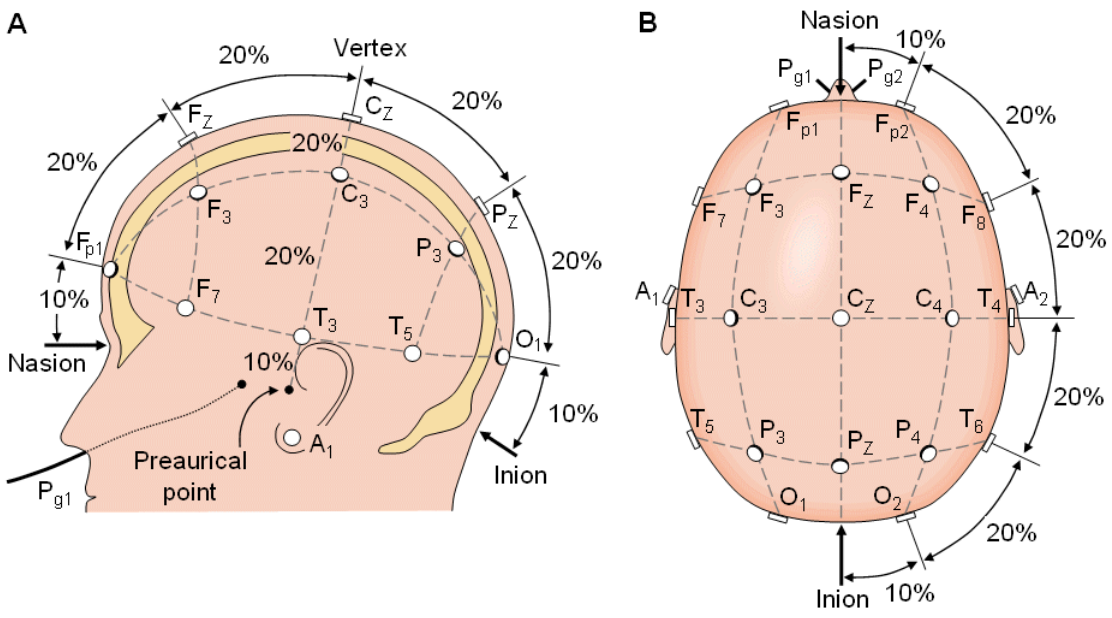

Figure 2.2: The International 10-20 system. This standard positioning system for EEG electrodes is widely used to ensure good reproducibility between experiments and across subjects. A View of the left side and B of the top of the head. Letters refer to the different brain regions (F: frontal, C: central, T: temporal, P: parietal and O: occipital). From [27].

from a specific location in the brain, but rather from many different regions at once. Different approaches to disentangling the different sources of brain activity behind an EEG signal have been proposed, one of which uses spatial filters (see Section 2.1.4). In contrast, given that it measures electrical current, EEG offers a high temporal resolution that typically reaches hundreds of Hz in modern systems.

Given that EEG measures electrical potential, at least two contact points are necessary to pick up a signal, one of which is used as a reference. Different conventions have been adopted to provide a standardized way of measuring EEG. The standard most often used, called the International 10-20 system [27] (see Figure 2.2), consists of positioning the electrodes in 10% and 20% marks along the lines linking the front (the nasion, Nz) and the back of the head (the inion, Iz), and the one linking the two sides of the head (A1 on the left and A2 on the right). This ensures good reproducibility between recordings on a same subject, and across subjects who have different head sizes and shapes. Extensions such as the 10-10 system (increments of 10% only) provide a higher number of recording sites.

To record these electrical potentials, electrodes are positioned directly on the scalp, often with conductive gel, paste or saline solution to reduce impedance where the sensors connect with the skin [28]. Since EEG requires the amplification of very low amplitude electrical signals, it can be strongly perturbed by electromagnetic noise such as line interference (60 Hz in North America and 50 Hz in Europe) and other sources of electrical noise such as appliances and electronics. EEG is also highly sensitive to unwanted signals called *artefacts* that pollute the real EEG signals and that can arise

from other physiological processes (heart beating, eye movement, muscle movement or even tongue movement) or from displacement of the sensors relative to the skin.

There is significant evidence that basic information about cognitive functions is encoded in the frequency components of the EEG signal [29–31]. Indeed, a few frequency bands are often used to characterize brain states: the δ (0-4 Hz), θ (4-8 Hz), α (8-12 Hz), β (12-30 Hz) and γ (30-100 Hz) bands are each correlated with specific mental states or processes. More details on the different EEG paradigms that can induce oscillatory processes are introduced in Section 2.1.2.

Functional Near-Infrared Spectroscopy

Whereas EEG directly measures neuronal activity via electrical potentials, NIRS relies on the changes in the optical properties of blood to indirectly measure neuronal activity [32]. Whenever a group of neurons is activated, more oxygen is required in the local tissues to sustain the increased energy consumption. Oxygen thus needs to be carried in the bloodstream from the lungs to the brain. To do so, the human body uses hemoglobin, a molecule that exists in two states: oxygenated (Oxy-hemoglobin) and deoxygenated hemoglobin (Deoxy-hemoglobin). Simply put, a deoxygenated hemoglobin molecule becomes oxygenated when four molecules of oxygen attach to it. Coincidentally, this change of state also has an effect on the optical properties of the hemoglobin molecule, which for this reason is referred to as a *chromophore*. Indeed, the absorption spectrum of HbR differs from that of HbO, especially in the near-infrared (700-900 nm) range. By monitoring this change in light absorption, it thus becomes possible to gain insights about the variations in local blood oxygenation, and consequently evaluate neuronal activity.

This is done using two types of probes called optodes: sources and detectors. Sources, typically LEDs or lasers, shine light of at least two different wavelengths in the near-infrared spectrum (for example, 760 and 880 nm) through the skin. Detectors, in turn, are built using photodiodes that capture light and convert it into an electrical signal. By positioning a detector on the scalp approximately 3 cm from a source, some of the photons emitted by the source will be reflected by tissues all the way to a neighbouring detector, following a banana-shaped trajectory (see Figure 2.3) [33].

The intensity of these measured photons in the two frequencies can then be interpreted using a mathematical relationship between the incident light intensity I_λ and the transmitted light intensity $I_{0,\lambda}$, called the Modified Beer-Lambert Law (MBLL) [32]:

$$A_\lambda = log\frac{I_{0,\lambda}}{I_\lambda} = (\alpha_{\lambda,HbR} \times [HbR] + \alpha_{\lambda,HbO} \times [HbO]) \times B_\lambda \times L + G, \tag{2.1}$$

where A_λ is the optical density for wavelength λ. The multiplicative constants in front of the concentration of Deoxy-hemoglobin (HbR) and Oxy-hemoglobin (HbO) are $\alpha_{\lambda,HbR}$ and $\alpha_{\lambda,HbO}$, the specific extinction coefficients of each chromophore at wavelength λ. L is the length between the source and

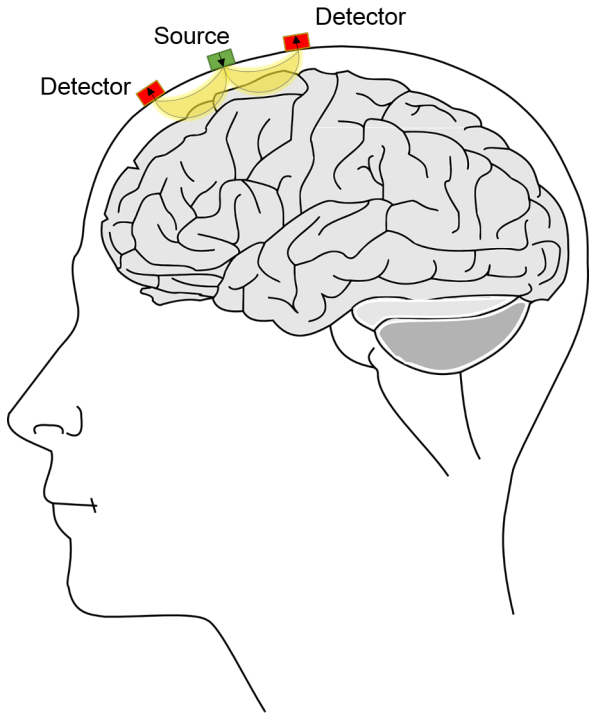

Figure 2.3: Illustration of NIRS signal generation. Light in the near-infrared spectrum is shined from a source through the scalp. Some photons, following a banana-shaped trajectory, will be reflected all the way to a neighbouring detector. The attenuation of the light intensity depends on the medium the photons are reflected through, which include parts of the cortex.

the detector, B_λ is the age-dependent *differential path length factor*, and G represents the signal losses caused by scattering. Focusing on the difference in Optical Density (OD) A_λ from one instant to another rather than on OD's instantaneous absolute value makes it possible to remove the constant G and thus obtain a system of equations linking changes in incident and transmitted light intensity to changes in chromophore concentration:

$$A/BL = \alpha C \tag{2.2}$$

$$C = \alpha^{-1} \times A/B \times L, \tag{2.3}$$

where $A = \begin{pmatrix} \Delta A_{\lambda_1} \\ \Delta A_{\lambda_2} \end{pmatrix}$, $\alpha = \begin{pmatrix} \alpha_{\lambda_1,HbR} & \alpha_{\lambda_1,HbO} \\ \alpha_{\lambda_2,HbR} & \alpha_{\lambda_2,HbO} \end{pmatrix}$ and $C = \begin{pmatrix} \Delta[HbR] & \Delta[HbO] \end{pmatrix}$.

The relative changes in HbO and HbR develop over approximately six seconds, a delay caused by the slow hemodynamic response function linking neuronal activity and blood flow phenomena [9]. Figure 2.4 shows the typical pattern observed in NIRS when a brain region activates. Once a cognitive task such as motor imagery starts, it takes a few seconds before a marked increase in [HbO] and a slight decrease in [HbR] can be noticed. After a plateau corresponding to the length of the task, these concentrations go back to their pre-task value another few seconds after the task ends. As a proxy for the total local

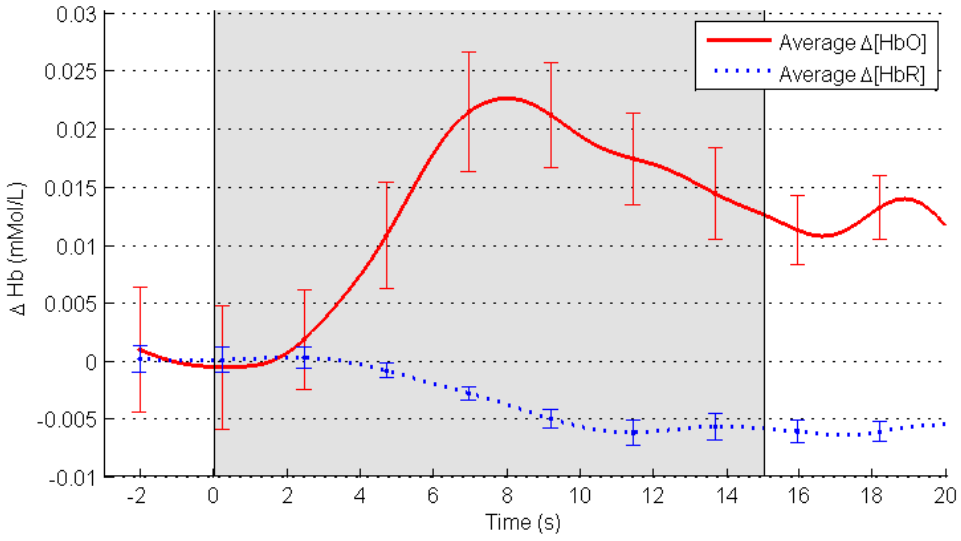

Figure 2.4: Dynamics of the HbO and HbR concentration signals induced by a word generation task. The shaded area represents the time during which the task was executed. The two lines represent the average HbO and HbR concentrations over many repetitions of the task. The error bars represent the standard error. (Data from S03, on channel S2-D5 over the left latero-central region.)

blood flow, the total hemoglobin concentration variation $\Delta[\text{HbT}] = \Delta[\text{HbO}] + \Delta[\text{HbR}]$ is also often used in NIRS analysis.

As opposed to EEG, the signals measured with NIRS represent the activity of a very localized portion of the brain (in the cm range) [34]. For this reason, NIRS is considered to offer good spatial resolution. However, due to the slow changes in NIRS signals and the long delay introduced by the hemodynamic response, NIRS is also considered to have relatively low temporal resolution, especially when compared to EEG. Moreover, unlike EEG, there is no standard way of positioning the optodes, but rather studies tend to focus on specific regions of interest generally corresponding to brain regions that are expected to be activated during the task being studied.

As is the case with EEG, NIRS suffers from physiological and movement artefacts, as well as ambient and instrumental noise [33]. Typically, heart pulse, respiration artefacts and low frequency drifts due to instrumentation have to be filtered out from the NIRS signals before analysis. Furthermore, signal quality can vary strongly based on the hair and skin types of an individual: denser and darker hair or skin will typically absorb more light, and thus decrease the number of photons that can reach the detector [103]. The same effect is caused by skull and skin thickness.

2.1.2 BCI paradigms

EEG-BCIs usually rely on one of six main groups of neural response patterns: sensorimotor activity, event-related potentials, steady-state evoked potentials, response to mental tasks, slow cortical potentials and other power changes [24, 44]. More details about these patterns are given below.

Sensorimotor activity includes neural response patterns that arise with motor execution (ME) or imagery (MI). Specifically, frontal and parietal cortices exhibit rhythmic activity in the 8-12 Hz and 13-30 Hz ranges, respectively called μ and β rhythms, when no sensorimotor activity is ongoing. Whenever a voluntary movement is triggered or imagined, however, these rhythms fade out, a phenomenon termed event-related desynchronisation (ERD). Similarly, once the movement is over, these rhythms emerge again and produce an event-related synchronisation (ERS). Another type of sensorimotor neural response pattern is a low-frequency, bilateral potential called movement-related potential, that arises around 1-1.5 s before a movement occurs.

Event-related potentials (ERP), in turn, are positive or negative deflections in the EEG signal caused by specific stimuli. The most preponderant ERP in the BCI literature is the P300, a positive deflection in the electrical activity of the parietal cortex occurring around 300 ms after the presentation of a rare stimulus, such as a visual, auditory or tactile cue. The error potential (ErrP) is another type of ERP that is measured when an individual recognizes an error that was made [104]. Although still not strictly characterised, the ErrP was found to consist of two successive deflections of the EEG: a negative one in the fronto-central region 50 to 100 ms after the onset, and a positive one over the fronto-parietal region around 200-400 ms after the onset [105, 106].

Steady-state evoked potentials (SSEP) are elicited by oscillatory stimuli with a constant frequency component. In response to this rhythmic stimulation, the sensory cortex exhibits phase-locked spectral activity with the same frequency as the stimulus [35]. For BCI applications, they are often produced with visual stimuli and recorded over the visual cortex, and therefore called steady-state visually evoked potentials (SSVEP). However, somatosensory and auditory stimuli are also capable of eliciting this kind of neural response pattern (SSSEP and SSAEP, respectively) [35].

Mental tasks are also known to elicit particular brain activity patterns. For example, mental calculation, mental rotation, covert spatial attention and selective sensation all produce different activation patterns that recruit different brain regions and follow different time courses [36–38]. Motor imagery, although described more precisely with neural response patterns such as ERD/ERS, is another example of a well-studied mental task. When volitionally evoked, reaching certain mental states such as a concentrated or relaxed state can also be considered as mental tasks. Since different mental tasks can produce significantly different activation patterns, we consider that each one of them produces a different neural response pattern. Although variations on a single task can be classified, for example motor imagery of the right hand versus the left hand, the foot or the tongue [107], we only consider mental tasks that involve different types of brain activity as producing different neural response patterns.

Slow cortical potentials (SCPs), in turn, are slow oscillations in the EEG arising from cortical polarization. One can consciously control them through adequate training [108].

Finally, **power changes** in other frequency bands or brain regions can be used to infer a user's affective states [44]. As seen in Section 2.1.8, this is important for so-called passive BCIs.

On the other hand, the neural response patterns used in **neurophysiological modalities based on hemodynamic brain activity** are generally not as diverse. For instance, in NIRS and fMRI, a localized increase in blood flow and oxygenation, called the hemodynamic response function (HRF), can be monitored to detect the activation of certain brain regions [32]. Since this activation varies in strength and spatial distribution with the type of mental task being performed [109], most hemodynamic BCIs aim at classifying mental states that are expected to induce different spatio-temporal patterns [11]. However, the hemodynamic response can differ substantially between two individuals or from one day to the other for a single individual [110]. Due to this high variability, hemodynamic BCIs usually rely on a learning algorithm that learns to distinguish between the studied states, instead of relying on the specific known features and characteristics of the signals [11, 111–113].

In this work, we focus on mental task-induced brain activity patterns from the perspective of both EEG and NIRS. Additional information on the use of mental tasks in unimodal and multimodal contexts is presented in Section 2.3 through a quick survey of the pertinent literature.

2.1.3 Artefact processing

Processing of the recorded neurophysiological signal(s) is done following three general steps: artefact processing, feature generation and feature translation. When designing a hybrid BCI, the integration of the different signals and modalities can be done at any of the two latter steps (see Section 2.1.7). In artefact processing, the signals are cleaned of the components that do not convey desired information and arise from other physiological processes (blood circulation, respiration, heart movement, eye blinks, etc.) or physical events (relative movement between the user and the modality) that pollute the signal of interest [39]. Artefacts are usually either rejected, corrected, or ignored. To do so, different types of filters can be used: for example, notch filters are often used to remove line interference in EEG, and lowpass filters to remove high frequency noise caused by muscle movement. Similarly, the NIRS signals are typically cleaned of the heart pulse signal and respiration-induced oscillation with some lowpass and highpass filters. Otherwise, studies have traditionally relied on visual analysis of the signals by an expert. However, real-time applications, such as BCIs, must use other techniques which can be automatically and reliably implemented in an online system [39]. The interested reader is referred to [40] and [41] for surveys of artefact handling techniques in EEG and NIRS, respectively.

2.1.4 Feature generation

Feature generation aims at extracting relevant features from the raw signals and can be divided into three subcomponents: preprocessing, feature extraction and feature selection [24]. To start with, preprocessing techniques are used to enhance the signals and to isolate some of their components, such as a range of frequencies or a spatial distribution, usually with different types of filters. Following artefact handling procedures, spatial filters are often used to enhance the EEG signals in particular. For example, the surface Laplacian consists of subtracting the signals from surrounding electrodes to get a more locally focused signal [114]. More advanced techniques, such as the supervised Common Spatial Patterns (CSP) filter, rather relies on maximizing the ratio of variance between linear combinations of electrodes for the two classes present in the data [115]. CSP is also great for dimensionality reduction by compressing most of the information in one or two components instead of multi-dimensional signals.

Next, feature extraction techniques can be used to extract precise information from the preprocessed signals. These so-called features are thought to indicate more or less reliably the presence or absence of a phenomenon in the user's brain activity (see Section 2.1.2). Examples of features extracted in the literature are discussed in Section 2.2.4.

Finally, in some cases, the number of possibly relevant features might be considerably large; it is therefore necessary to select the most informative, which can be done using dedicated algorithms. With multimodal systems, the amount of data that is recorded can be extensive, particularly if multiple channels are used per modality, and several features are extracted per channel. As such, feature selection becomes a necessary step in order to avoid the so-called "curse of dimensionality" during classifier training [116]. Generally, this also helps to simplify the subsequent steps of the BCI and to reduce its computational burden.

2.1.5 Feature translation

Feature translation is the last processing step, and consists of making sense of the previously extracted features with the use of a pattern recognition algorithm such as a classifier or with a set of pre-defined rules [42]. Depending on the configuration of the system, many such classifiers can operate in parallel, in which case their respective decisions are combined and fused by a given set of rules or a meta-classifier (see Section 2.1.7). We focus our discussion of feature translation techniques on classification, as this is the most often used procedure in brain-computer interfacing.

Classification framework

A classifier operates on a dataset D that contains n data points $x \in \mathbb{R}^d$, where d is the fixed dimensionality of a data point. Each data point in D is described by a label y that represents the class of

that point. The labels thus take discrete values such as -1 and 1 in the case of a binary classification problem. The dimensions of D can vary a lot from one problem to another, but to avoid problems such as the curse of dimensionality, it is good practice to collect data such that $n > d$.

The goal of classification is to learn a function f_θ that can map a new, unseen example x' to its correct class y'. In the context of a BCI, this means a classifier needs to identify the type of mental state an individual is in given d different features reflecting that person's neurophysiological processes. Many different algorithms can be used for that purpose (see [42] for a review).

To obtain an optimal function f_θ, a classifier has to learn the structure of the data. To do so, it is provided with a *training set* D_n, a subset of the whole dataset D. During the learning phase, the classifier adjusts its parameters θ in order to minimize the classification loss $L(f_\theta(x), y)$ on D_n. This loss function penalizes wrong classifications and thus provides a measure of how correct the classifier is when making decisions about the examples contained in the training set. During the subsequent validation phase, the generalization performance of the trained algorithm f_θ is evaluated on a second subset of D, the *validation set* D_m, which was never seen by the algorithm during training. This step is very important to ensure the classifier has learned to generalize, as opposed to remember data-specific information that is not relevant on new data (a phenomenon called *overfitting*). Finally, in addition to the parameters θ that we want to optimize, classification algorithms have hyper-parameters that define and control their complexity. Different types of algorithms will also have different hyper-parameters (see 2.1.5 for examples). In the case where we want to compare different algorithms with different hyperparameters, we need to further partition our training set D to reserve data on which these different configurations can be compared. It is then possible to select the best performing algorithm and use it on new data.

A major concern when training classifiers is the dimensionality of the data. Indeed, when the number of features is relatively larger than the number of data points in the training set ($d > n$), many algorithms will perform poorly and will not be able to generalize well. This is often the case when dealing with neurophysiological data, where many data points are collected (e.g., from each electrode of an EEG setup), but where the number of repetitions of each class is low. This issue can be mitigated by selecting important features, reducing dimensionality with techniques such as Principal Components Analysis (PCA) and Common Spatial Patterns (CSP), or by selecting classification algorithms that are less prone to overfitting.

Linear classifiers

While many types of classifiers have been proposed, most hBCI implementations in the literature tend to use the same linear classifiers [14], as opposed to non-linear classifiers such as neural networks, decision trees, k-nearest neighbours, etc. Linear classifiers work by finding a d-dimensional hyperplane that best separates the n data points into their respective classes y. A typical linear classifier can be formulated

as:

$$f(x) = sign(\langle \boldsymbol{w}, \boldsymbol{x} \rangle + b), \tag{2.4}$$

where the parameters are $\theta = \{\boldsymbol{w}, b\}$, with $\boldsymbol{w} \in \mathbb{R}^d$ the weight vector and $b \in \mathbb{R}$ the bias. The $sign()$ function returns 1 if the sign of its argument is positive, or -1 is it's negative, which would be used in a binary classification problem. The principal difference between linear classifiers is the way the hyperplane is found (i.e., the training procedure); the mathematical formulation otherwise stays essentially the same.

The Support Vector Machine (SVM) is a very popular linear classifier (that can also be generalized to a non-linear classifier) based on a geometrical view of the decision boundary [43]. SVMs have been used successfully in an extensive number of different applications [117]. The SVM algorithm finds an hyperplane that maximizes the distance between the hyperplane itself and the closest training data points - a distance called the *margin* - while separating the data points according to their respective class (see Figure 2.5). The corresponding optimization problem can be written as:

$$\min_{\boldsymbol{w},\zeta} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{n} \zeta_i, \tag{2.5}$$

with respect to $(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle + b) y_i \geq 1 - \zeta_i$ and $\zeta_i \geq 0$. C is a regularization hyper-parameter that controls the tolerance against misclassification (a smaller C means more tolerance), while ζ_i are *slack variables* that are 0 if their corresponding data point is on or inside the appropriate margin, and otherwise will grow above 0 depending on how close they are to this boundary. The SVM training algorithm is thus designed to find the smallest combination of $\boldsymbol{w}$ and ζ_i such that the $\boldsymbol{x_i}$ are generally correctly classified. Thanks to the so-called *kernel trick*, extensions of the SVM algorithm allow the use of non-linear decision boundaries, providing higher performance when the data structure cannot be simply separated by an hyperplane.

Many other linear classifiers are widely used in the hBCI literature. For example, the Linear Discriminant Analysis (LDA) works under the assumption that each class is normally distributed, and thus aims at finding the hyperplane that maximizes the distance between the means of the two classes, while minimizing the interclass variance [42]. Another widely used linear classifier is the Logistic Regression, which works by using a non-linear sigmoid transformation on top of the linear activation (instead of the $sign$ function), allowing a binomial probabilistic interpretation.

Finally, once a decision concerning the user's current mental state has been made using a classifier or another decision-making algorithm, the output is fed into a control device, such as a display, prosthesis or robot, to name a few. This feedback can be visual, auditory or tactile, or not directly perceptible to the user like in some passive BCIs (see Section 2.2.2). For specific types of BCIs, particularly with visual displays, the feedback interface can also provide the stimuli necessary to elicit the brain activity patterns required for driving a BCI.

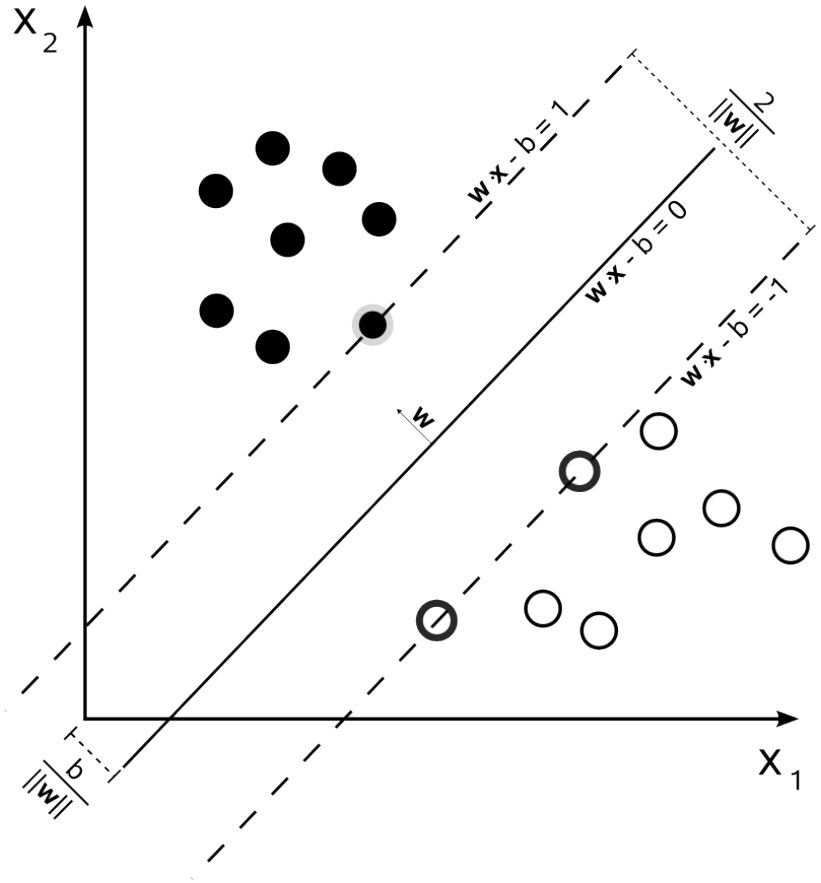

Figure 2.5: Hyperplane selection for a linear SVM. In this binary classification problem, two-dimensional data points from class 0 (in black) and class 1 (in white) can be perfectly separated by an hyperplane. That plane is chosen such as the margin $\frac{2}{||w||}$, that is the distance between the hyperplane and the closest data points, is maximized. Image taken from [118].

2.1.6 Performance metrics

The performance of a BCI is often evaluated based on its classification performance or its capacity to correctly transmit a certain amount of information in a given time. The classification accuracy, computed as the number of data points correctly classified divided by the total number of points in the dataset, is often used in the first case. In the second case, the Information Transfer Rate (ITR) and related metrics such as the Practical Bit Rate (PRB) are normally used. By taking into account the number of possible commands, the classification accuracy and the speed at which classification are carried out, these metrics provide a more complete perspective on the performance of the global system. A BCI that has a very high accuracy but that takes a long time to process each example will thus have a low ITR, making it more comparable to BCIs with lower accuracy, but allowing the user to input more commands more quickly. Finally, task-specific metrics can be used in some cases to provide more precise information about the use of the BCI in that context. For example, the similarity between the optimal path a BCI-controlled wheelchair could have taken to go from point A to B and the path taken by the user could be used

to show how the whole system behaved, rather than focusing on low-level classification of tasks that does not take into account the different strategies a user can use to perform a task adequately. More information about the performance metrics used in hBCI studies is given in Section 2.2.6.

2.1.7 Fusion levels

The strategy for integrating different signals can be categorized according to the level at which this fusion is done. Three such levels are typically recognized: data-level, feature-level, and decision-level fusion [44]. Data-level fusion is used when unprocessed signals carrying similar information are fused together. This avoids the loss of information when fusing the signals, but at the same time is more sensitive to noise and modality failure. An example of this type of fusion would be using the EEG signals coming from electrodes placed on two distinct regions of the head – i.e., the occipital and frontal regions – to compute information theoretic features, such as coherence.

Feature-level fusion occurs when the original signals are first processed individually before their resulting features are fused. This decreases the level of information detail that is available when fusing, but increases the robustness of the system. For instance, when trying to decode hand movement, extracting features from EEG and arm EMG separately and then feeding these features into a single classifier, would be considered feature-level fusion.

Decision-level fusion is the highest level at which signals can be integrated. Typically, this implies that different signals are individually processed and classified (thus producing a decision for each signal) before being fused. The fusion can then be aimed at controlling a single task, or the different interpreted signals can each control different tasks working together toward the same high-level task (termed semantic-level fusion). By the time the fusion is carried out, details present in the original raw signals have already been discarded. As a result, the system is even more robust to noise and possible failures. Systems combining a BCI with a non-physiological modality typically use this level of fusion: for example, an eye tracker monitoring gaze direction on a computer screen, in conjunction with an EEG-BCI used to select an item on the screen, would require decision-level fusion.

2.1.8 Subtypes of BCIs

In addition to its general definition, multiple BCI subtypes have been defined:

- Sequential and simultaneous (modality synchronisation);
- Active, reactive and passive (operation mode, similar to exogenous and endogenous);
- Dependent and independent;
- Synchronous and asynchronous;

- Discrete and continuous.

Modality synchronisation, or the way the different components of an hBCI interact together, is described as either sequential or simultaneous. In a *sequential hBCI*, the components are used one after the other, in such a way that the first one often acts as a switch allowing the user to activate or deactivate the second component. The first component can also act as a selector, allowing to switch between multiple possible choices or control interfaces [12]. In a *simultaneous hBCI*, on the other hand, every component can be purposefully controlled at the same time. This can be used to control two or more different parameters concurrently by mapping each modality or neural response pattern to one of them, or to increase the reliability of the BCI by providing multiple redundant inputs to control one parameter.

BCIs are also categorized according to their operation mode, that is the way the brain activity necessary for operating the interface is triggered [119]. *Active BCIs* rely on user-controlled, internally elicited brain activity patterns. For example, motor imagery tasks can be initiated and performed by a user without any external intervention. *Reactive BCIs*, on the other hand, rely on brain activity elicited by a given external stimulus, on which the user can voluntarily focus or not. The P300 event-related potential is one example in which the user's reaction to a stimulus generates distinguishable brain activity. The active and reactive modes are very similar to the endogenous and exogenous categorization reported in [96, 120], where *endogenous BCIs* rely on internal control exerted by the user, just as *active BCIs*, and where *exogenous BCIs* are driven by external stimuli, like *reactive BCIs*. Lastly, *passive BCIs* monitor patterns of arbitrary mental activity generated without the objective of volitional control. This is typically used to complement human-machine interaction by providing information on the user's current affective state. The identification of classification errors based on the recognition of the ErrP component has also been investigated.

BCIs have also been classified based on their dependence on peripheral or central nervous system control. *Dependent BCIs* rely on natural brain output pathways, such as peripheral nerves and muscles, to elicit the necessary activity for driving the system [1]. For instance, visual protocols such as the standard P300 speller require eye gaze direction to attend the stimuli, a process in which muscles play a central part. *Independent BCIs*, on the other hand, do not require any such peripheral functions, and are for instance well adapted to individuals who lost control of their muscles. Despite similarities in the categorisations, independent BCIs are not necessarily active (or endogenous). Indeed, BCIs based on covert visual attention such as some SSVEP [121] or P300 [122] interfaces were shown to work without eye gaze movement and thus do not necessitate muscle functions.

A BCI is *synchronous* if its inputs are analysed during predefined periods of time, initiated by the interface. An *asynchronous BCI*, on the other hand, is continuously analysing its inputs and can thus respond to commands at any given time [96]. The advantage of synchronous BCIs over asynchronous ones is that false positives cannot occur outside the time windows in which commands can be issued. While this may improve overall BCI accuracy, it limits the usability of the BCI as well as the information transfer

rate, since the user can only issue commands at specific periods of time. Conversely, asynchronous BCIs can be more flexible, but require robust decision-making in order to avoid high false positive rates when the user does not intend to issue commands.

Finally, a BCI can be characterised by its output type: as described in [123], *a discrete BCI* has few outputs that typically arise from a classification stage implying a choice between a few different possibilities, whereas a *continuous BCI* has outputs that can take any value in a specific range and is thus more suited to tasks such as cursor control.

2.1.9 Rigourous definition of a BCI

BCIs have been defined in multiple ways since the field's early days. However, as emphasised by a survey undertaken by members of the BCI community in 2010, the definition of a BCI is still not unanimous: around 60% of respondents agreed that a passive BCI (defined in Section 2.1.8) should be considered a BCI, and 50% thought that providing feedback to the user when a goal was successfully reached was necessary [124]. Wolpaw's definition [1, 125], although commonly accepted and cited, has thus undergone many refinements over the years. For example, Pfurtscheller et al. [12] defined four conditions that a communication system must fulfil to indeed be considered a BCI:

1. The signals on which the system relies must be recorded directly from the brain;
2. At least one of the system's input brain signals must be voluntarily controllable by the user;
3. The system must process information in real-time;
4. The system must provide feedback to the user.

This definition therefore excludes studies in which the recorded brain activity is not targeted towards volitional control, but rather towards monitoring of arbitrary activity, so-called passive BCIs. Studies in which the data processing and analysis are done offline are set apart as well. Following this definition, a hybrid BCI would need at least one of its components to fulfil those four criteria [13]. In another definition, Zander et al. [126] also underlined the importance of brain activity and of real-time processing, but did not consider volitional control and user feedback necessary.

More recently, in the first edition of the *Brain-Computer Interfaces* journal [127], a broader definition was prescribed: "A brain-computer interface is a system that measures central nervous system (CNS) activity and converts it into artificial output that replaces, restores, enhances, supplements, or improves natural CNS output and thereby changes the ongoing interactions between the CNS and its external or internal environment" [2]. The focus of this definition is on the role of the communication system rather than on the way it is implemented; it suggests brain activity must be used for generating an artificial output that somehow influences the world, including the user himself. This paves the way for a

generalization of brain-computer interfaces to brain-monitoring devices and systems that do not instantly provide feedback to the user.

2.1.10 Definition of a hybrid BCI

Simply described in [128] as a "combination of different signals including at least one BCI channel", hBCIs have been given a wide definition that allows systems relying on multiple neural response patterns or diverse modalities to be referred to as such. Allison et al. [128] also proposed to categorize hBCIs according to the type of combination used, namely: *pure hBCIs*, which combine two or more conventional BCI approaches, *physiological hBCIs* which combine a BCI with another physiological signal, or *mixed hBCIs*, which combine a BCI with a non-physiological input [128].

Similarly, in [13], six types of hBCIs were identified, based on:

1. Two different EEG BCIs;
2. EEG and a non-EEG BCI;
3. EEG BCI and another biosignal;
4. EEG BCI and EEG-based monitoring;
5. EEG BCI and other signals;
6. EEG BCI, EEG monitoring and other biosignals.

If a passive BCI is recognized as a BCI, as in Wolpaw's broader definition [2], the addition of EEG-based monitoring (which could be extended to monitoring with other physiological modalities) becomes redundant. Therefore, although this categorization seems more exhaustive, we refine it by including passive BCIs and generalizing to non-EEG-based BCIs, as depicted in Figure 2.6.

A physiological hBCI thus requires that one of its components be a non-neurophysiological modality, independently of whether it is used for active or passive control of the interface. Pure hBCIs can now refer to any combinations of BCIs as defined above. Additionally, as suggested in [22], hBCIs may be categorized according to the way their multiple inputs are combined: each input is mapped to a separate application, all the inputs are fused and weighted and map to a single output [129], or an additional monitoring module decides which input should be used as a control signal, based on evaluation of the signal quality.

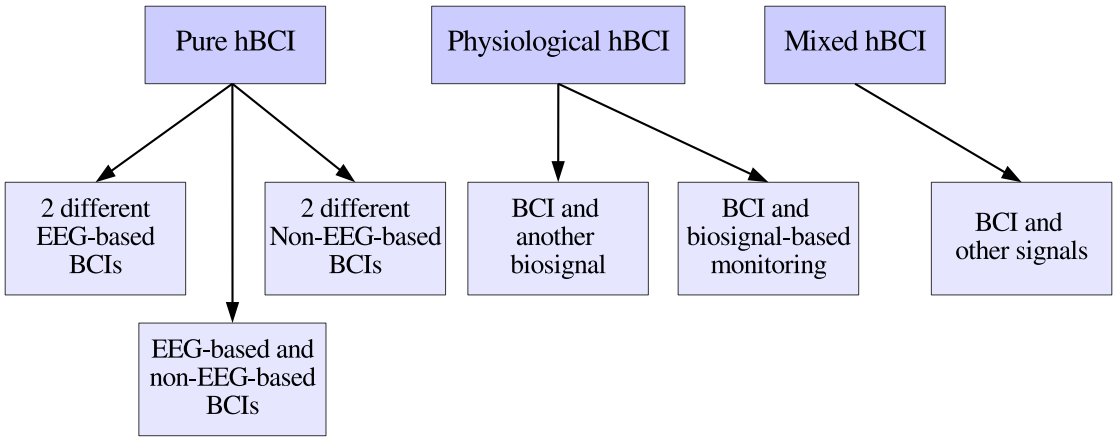

Figure 2.6: Taxonomy of hybrid BCIs.

2.2 Results from a systematic review of the hybrid Brain-Computer Interfacing literature

Three major databases covering the fields of science and engineering were queried. From these and other sources, 55 journal articles from 2008 to November 2014 were selected and analysed. Twenty-two items were investigated to offer a complete perspective on the current state of hBCI research, covering the study rationale, experimental protocol, signal processing methodology and system evaluation. However, only the most relevant sections are reported here, for the sake of conciseness. The detailed methodology used to select the reviewed articles is fully described in Appendix A.

2.2.1 Hybridisation paradigms

Two main types of hybridisation can be distinguished in the selected articles: (1) modality-based hybridisation, when two or more modalities are used together (that is, mixed or physiological hBCIs), with at least one of them probing neurophysiological functions (EEG, NIRS or fTCD), and (2) neural response pattern-based hybridisation, when different neural response patterns are used in conjunction, with EEG alone (pure hBCI). The different modality combinations used by the selected articles, as well as the corresponding hBCI types, are presented in Table 2.1.

The vast majority of studies used EEG in their protocol, while only a few used NIRS. Among these, 31 were pure EEG hBCIs and thus were based on the hybridisation of two or three [155] neural response patterns recognisable in EEG. Similarly, four were pure EEG-NIRS hBCIs [17–19, 168]. The remaining

Table 2.1: Hybridisation paradigms of the selected hBCI designs, following the used modalities and neural response patterns. Refer to Section 2.1 for the definition of acronyms used here.

Modalities (# of articles)	hBCI type	Articles
EEG only (31)	SSVEP + SSSEP	[130]
	MI + SSVEP	[12, 128, 131–136]
	MI + P300	[137–143]
	MI + selective sensation	[144]
	P300 + ErrP	[145, 146]
	Audio + tactile P300	[147]
	P300 + SSVEP	[148–154]
	MI + P300 + SSVEP	[155]
	Multiple mental imagery tasks	[36, 51, 53]
EEG + electrophysiological modality (7)	P300 + EOG	[156]
	MI + EOG	[157]
	MI + P300 + EOG	[158]
	MI/ME + EMG	[129, 159]
	ME-based monitoring + EMG	[160]
	ME-based monitoring + EMG + Inertial sensors	[161]
EEG + external signal (9)	MI + eye-tracker (gaze position)	[126, 162, 163]
	EEG-based monitoring + eye tracker (pupil dilation)	[164]
	Mental task + eye tracker (position)	[165]
	SSVEP + EEG-based monitoring + keyboard	[21]
	MI/ME + game controller	[22, 166]
	MI + analog position sensor	[167]
NIRS + EEG (4)	NIRS-EEG-based MI/ME	[17, 19]
	NIRS-EEG-based SSVEP	[168]
	NIRS-based mental imagery + EEG-based ME	[18]
NIRS + physiological signals (3)	NIRS-physiological signals-based ME	[169]
	NIRS-physiological signals-based mental imagery	[57]
	NIRS-fTCD-based mental imagery	[65]
NIRS + EEG + physiological signals (1)	NIRS-EEG-physiological signals-based monitoring	[170]

articles relied rather on a modality-based hybridisation paradigm. These combinations of two or more modalities can be divided into four different subtypes: EEG and electrophysiological modalities, EEG and external signals, NIRS and physiological signals, and NIRS, EEG and physiological signals. In all of these cases, the system relies on a neural response pattern that can be identified in EEG and/or NIRS, while the added modalities either contribute a traditional physical movement-based input or instead contribute supplementary data on the subject's physiological state, without relying on a specific neural response pattern. Electrophysiological signals like EMG and EOG and input devices were usually employed with EEG, whereas physiological signals (SCR, respiration, skin temperature, blood pressure, fTCD) were always used at least with NIRS. In total, 14 different modalities were used in the reviewed articles.

The ubiquity of EEG in the reviewed studies is to be expected given the actual state of BCI research. Indeed, EEG has been the modality of choice since the field started thanks to its well-studied features and practicality [171]. For example, in comparison to EEG, NIRS is less suited for near immediate control of a BCI due to the delay introduced by the cerebral hemodynamic response, which is of a few seconds [9]. Nonetheless, a more complete understanding of the underlying physiological phenomena and a technical refinement of the current devices might eventually compensate for this shortcoming of NIRS and other hemodynamic modalities. Specific features might also allow for faster NIRS signal classification [168].

Non-neurophysiological modalities can complement EEG and NIRS in two ways. First, physiological measures such as SCR, pupil dilation, respiration and pulse can be analysed as a manifestation of conscious or unconscious cerebral activity through the autonomic nervous system, and thus help discriminate mental activities such as mental singing [57]. It also offers a different perspective for emotion and stress level assessment, such as in passive BCIs [23]. Second, physiological and external signals such as EMG, EOG, eye trackers and game controllers can be used as conscious inputs that stand on their own, and allow much more reliable and precise control of an active interface. In both cases, however, using additional non-neurophysiological signals might prove useless in patients suffering from spinal chord lesions or neurological conditions, who may lose part of their sympathetic response [172] (a pure hBCI would then be the only hybrid usable).

The most frequent hybridisation paradigm in the reviewed articles is a combination of motor imagery and a visually evoked signal feature such as P300 or SSVEP. The number of currently established neural response patterns being low, this combination is a good way of producing distinguishable signals, while improving the robustness of the BCI. Indeed, these brain patterns exhibit different spatial and temporal signatures, and also arise from different operation modes (namely, active and reactive BCIs). Achieving this complementarity was a goal in [36, 51, 53], in which multiple mental tasks were tested in order to identify pairs of tasks that are easily distinguishable. Thus, the choice of mental tasks can be adapted to each individual user so the BCI performance is the highest possible. Moreover, this type of pure unimodal hybridisation paradigm is also attractive because it requires less hardware and consequently is easier to install and use than systems based on multiple modalities.

2.2.2 Operation mode

The type of BCI operation (active, reactive or passive) is shown in Table 2.2 for each selected article. As mentioned in Section 2.1.9, BCIs can be classified according to how the neural response patterns necessary to the system's functioning are generated. Taking into account the effect mental activity can have on the sympathetic nervous system, for example on heart rate [173], mental processes induce changes in the signals of both physiological and neurophysiological modalities. This is the motivation behind procedures such as biofeedback, which aim at gaining control over normally unconscious processes through feedback on physiological signals. However, this is obviously not the case for non-physiological inputs, such as a game controller. Therefore, in the case of hybrid systems, only the system parts that rely on a physiological modality can be classified as active, reactive or passive.

Reactive systems were usually composed of a combination of P300 and SSEP. In the case of passive systems, different approaches were reported, including the recording of passive reactions to a series of presented stimuli; in [160] and [161], which both reported tremor management systems, subjects performed movements that triggered the use of a prosthesis, but without the explicit purpose of voluntary controlling of the BCI. This makes them in our opinion passive BCIs.

Table 2.2: Operation mode of the selected hBCI designs.

BCI type (# of articles)	Articles
Active (19)	[17–19, 22, 36, 51, 53, 57, 65, 126, 129, 157, 159, 162, 163, 165–167, 169]
Reactive (11)	[130, 147–154, 156, 168]
Passive (4)	[160, 161, 164, 170]
Active-Reactive (19)	[12, 21, 128, 131–144, 155, 158]
Reactive-Passive (2)	[145, 146]

Two composite operation modes (active-reactive and reactive-passive) were found in the reviewed articles. In [21], the power of the subjects' alpha waves was used in conjunction with SSVEP to control the oscillatory displacement of a game element. The control of the alpha waves can thus be seen as an active component, since the subject consciously modulates it to improve his in-game performance. Most of the other active-reactive systems simply combined two neural response patterns that happen to involve either active or reactive operation modes. Finally, reactive-passive systems made use of the ErrP to enable command confirmation in a P300 speller [145, 146].

Active BCIs are especially practical for synchronous control. Indeed, active BCIs do not rely on any external stimulus: it is therefore more natural for a user to perform a command when desired since the initiation and performance of the necessary mental activity are completely internal; this may also help reduce fatigue in the user. Furthermore, reactive BCIs found in the reviewed articles used stimuli that were not aimed at modulating processes other than neurophysiological ones. Physiological hBCIs thus could not benefit from reactive paradigms as much as pure hBCIs.

Passive BCIs, on the other hand, aim at different applications where direct control of the interface is not required, and are still in their early stages [23], thus are not yet widely used. As compared to active and reactive BCIs, passive BCIs by definition do not induce fatigue or workload, but rather typically attempt to measure such states to adapt the behaviour of a system to the user. As such, they could be a useful addition to complex, fatigue-inducing hBCIs based on active and/or reactive operation modes. It is interesting to note also how the definition of "passive" can be challenged by the designs found in [160, 161, 164]. In these studies, the neural response patterns that are used are typically found in reactive or active BCIs: in [164], the presentation of images at high speed elicits an event-related potential such as P300, whereas in [160, 161], the initiation of a movement of the arm is picked up to activate a stimulation system. However, since in these cases the user is not voluntarily producing the neural response patterns, but rather is producing them as part of normal, undirected behaviour, we believe these are examples of passive BCIs.

2.2.3 Application

Eight categories of applications can be distinguished from the selected articles, as shown in Table 2.3. Only the most important application for each study is reported. Most articles did not aim for a specific

Table 2.3: Reported application of the selected hBCI designs.

Application (# of articles)	Articles
General improvement of hBCI or no specific application (16)	[17–19, 36, 51, 53, 57, 65, 128, 130, 131, 133, 134, 144, 157, 168]
Assistive technology and rehabilitation (13)	[12, 22, 129, 132, 143, 146, 147, 152, 159–161, 167, 169]
Spelling (10)	[126, 145, 148–150, 153, 154, 156, 162, 163]
Navigation (8)	[135, 136, 138, 140, 151, 155, 158, 165]
GUI control (4)	[137, 139, 141, 142]
Video gaming (2)	[21, 166]
Affective monitoring (1)	[170]
Decision aid system (1)	[164]

application, and rather proposed general improvements on hBCIs. Indeed, 16 articles presented a new hBCI paradigm or signal processing methodology without specifying an application for the system.

The category of assistive technology and rehabilitation covers different subfields: prosthesis control [12, 132, 143], assistive technology for managing muscular fatigue [22, 129], tremor management using a functional electrical stimulation (FES) device [160, 161], rehabilitation tools [167, 169], consciousness monitoring [152], and spelling targeted at or evaluated on disabled individuals [146, 147, 159].

Hybrid spelling applications improved on two different initial concepts for spellers: the P300 speller, and an eye tracker speller. In the first case, several improvements were proposed to the standard P300 speller: adding SSVEP flicker to the screen to know when the user is looking at it [148], using EOG to know which half of the screen the user is looking at [156], using the halting in SSVEP flickering of the letters to elicit a P300 [149, 153], combining SSVEP flickering of the letters with the usual P300 paradigm [150, 154] and adding a passive validation of a character selection using the ErrP [145] (although classified as assistive technology, a similar concept was explored in [146]). In the second case, [126], [162] and [163] proposed to use motor imagery to select a character indicated by an eye tracker-controlled cursor.

Navigation applications were applied to virtual environments [138], virtual or real wheelchair control [135, 136, 140, 151, 158] and to remote robot control [155, 165]. The control of a graphical user interface, in turn, was attempted across a few studies. First, cursor control was implemented in [137] with P300 and MI; second, an enhanced implementation using an additional P300 button for selection was developed in [139]; third, a web browser was developed using the previous cursor methodology in [141]; finally, an email client using the same control concepts was presented in [142]. The same design was applied for wheelchair control in [140].

Two articles explored video gaming applications. In [21], a 2D video game based on alpha waves neurofeedback and SSVEP was developed in which the player controls an amoeba trying to eat bacteria. In [166], the player controlled the jumping of a penguin in a race using motor imagery or execution.

Finally, two studies using passive systems aimed at different applications: one article [170] presented an attempt to infer subjects' affective states using EEG and physiological signals, while the other [164]

reported a passive system to identify images of interest presented to the subjects by analysing EEG and pupil dilation activity.

Promising applications for BCIs in general, both in medical and non-medical fields, were highlighted in previous reviews [3, 174–177]. Most of the medical applications mentioned in these reviews were the focus of some of the selected articles: assistive technologies such as spellers, navigation, user interface control, FES device control for grasping and rehabilitation. However, in the case of non-medical applications, there is still room for innovation: for instance, user-state monitoring, as made possible by passive BCIs, was mentioned as a promising application for BCIs in [3, 175] but was the focus of only one of the selected articles [170]. This type of application is better suited to non-medical users since it is not in competition with standard user interface technologies. Indeed, active and reactive BCIs cannot achieve the same communication bandwidth as a mouse or keyboard, at least as of now, and thus healthy users are less likely to adopt these as standard input devices. Passive BCIs rather provide an additional input related to a user's affect, mental workload, fatigue or attention. Moreover, hybrid BCIs can readily benefit from additional modalities typically used in affect recognition such as SCR and PPG. Additionally, gaming and entertainment applications are highly promising, as these industries are often early adopters of new technology [3]. Hybrid BCIs are particularly relevant for this type of application: mixed hBCIs combining traditional controllers and brain-driven inputs can naturally enhance a gamer's experience without suffering from the bandwidth bottleneck of BCIs used for direct control. Other applications that were not mentioned in the selected articles include applications in training and education, cognitive enhancement, neuroergonomics and neurofeedback [3].

2.2.4 Commonly used features

Table 2.4 shows the different features used in the articles, based on the modalities and neural response patterns that were used. In studies relying on EEG, features were extracted for six different types of neural response patterns: ERP, SSVEP/SSSEP, MI/ME, alpha waves, mental tasks-, and passive brain activity-induced patterns. For P300, the articles mainly used amplitude values of data points spanning the 600-ms window after a rarely occurring stimulus. Nearly half of them averaged a few of those windows together in order to get their feature vectors, whereas the other half only concatenated multiple windows together to form them, as was the case for the two articles involving ErrP [145, 146]. In [155], the so-called xDAWN unsupervised spatial filter was applied before the extraction of amplitude values, to maximise the signal-to-signal-plus-noise ratio [178]. Two articles explored different approaches: in [149], the features were the averaged spectral power across multiple epochs, combined with the inter-trial phase coherence, which was concurrently used for SSVEP detection; in [147], a 600-ms window was transformed in the frequency domain using the Synchrosqueezing Transform and then bandpassed between 0.1 and 40 Hz to keep the information related to the P300 response.

For SSEP, different features were extracted: band powers (or logarithm of the band powers) around the fundamental frequency and/or the first and second harmonics, variations on the band powers (Har-

monic Sum Decision [12], band power components extracted with Correlation Components Analysis (CCA) [150], band powers of Minimum Energy Combination [151], averaged spectral power with inter-trial phase coherence, which was used at the same time for P300 detection [149]); time points [132]; Harmonic Phase Coupling [131] and CCA correlation components [135, 136, 154, 155].

For motor tasks, most articles used (logarithmic) band powers with varying widths and durations or features extracted from the application of the Common Spatial Patterns (CSP) supervised filter. This process was optimised in [143] by first finding subject-specific time intervals and frequency ranges. Additionally, the coherence values between EEG and EMG were used as features in [160]. Only one article [18] used the average peak amplitude.

Finally, other EEG studies used the band power over alpha frequencies (for alpha waves monitoring [21]), logarithmic variance of CSP components or band powers extracted from CSP components (for mental task detection [36, 51, 53] and [165], respectively) and cross-correlation patterns between the different transformed channels (for reactions to affective stimuli) [170].

In NIRS, all articles used delta-concentration values obtained with the MBLL, with the exception of [57] which used position-averaged light intensity values. In [169], instead of using time points, the coefficients of a linear combination of the original channels were used as features. Finally, in [65] and [168], the slopes of Hb, HbO and HbT were employed as features.

Other modalities also had their specific features. For example, eye tracking was used to gather information on pupil size and movement through 22 different features in [164]; EOG was analysed using the amplitude of bipolar channels [157] or correlation components from CCA [158]; EMG was either averaged, or fed into a Hilbert transform [159, 161]; the slope of the mean, change from the start to end of trial and range of the amplitude values of fTCD signals were used in [65]; other physiological signals were usually employed as is (following their preprocessing) and concatenated.

Of particular interest in the case of hBCIs is the computation of features extracted from more than one neural response pattern or modality at a time. This could highlight the crosstalk between different physiological processes and thus improve the accuracy of the system. For example, as was mentioned above, in [160] the coherence between EEG and EMG was used to better detect the intention of movement in subjects with tremors; in [170], cross-correlation patterns between the Hilbert-transformed components extracted from the Empirical Mode Decomposition of all modalities were used as features; finally, in [36, 51, 53], the extracted features can be viewed as a mixing of the different neural response patterns (induced by mental tasks) at use, fused by the application of the CSP technique. As discussed in Section 2.1.7, other data-level fusion approaches could improve BCI performance while being of a particular physiological relevance.

2.2.5 Fusion level

The fusion level used in each selected article, as covered in Section 2.1.7, is shown in Table 2.5. Whereas most studies used decision-level fusion, only a few studies used data-level fusion: in [160], the specific feature used (cortico-muscular coherence) required the data fusion of both EEG and EMG; in [36, 51, 53, 144, 147, 149], the features were the same for each of the tasks and were thus directly extracted from the same EEG data. In [139, 141, 142], semantic-level fusion was used on top of the feature-level fusion for target selection and the decision-level fusion for cursor control in order to merge the two tasks together.

In addition to the distinction between decision- and semantic-level fusion, it is also possible to vary the way the different decisions are fused. In [129], two approaches were tested: a standard fusion where the two decisions were equally balanced, and a naive Bayes fusion where the weights of each decision were adapted. It is also worth mentioning the special fusion paradigm reported in [22], where the system did not use both modalities (EEG and joystick) concurrently, but rather switched between one and the other depending on their respective signal-to-noise ratio. This approach was also proposed in [25] as a general framework for hBCIs where brain-driven inputs are complementary to the other standard devices, and in which the system is able to switch from one modality to another.

Data-level fusion is a promising avenue for enhancing the performance of multimodal BCIs. For instance, neurovascular coupling as measured in joint electrophysiological and hemodynamic modalities [179] and phase-amplitude coupling between EEG and SCR [180], show both a theoretical physiological relevance and could bring additional information unimodal systems based on feature-level fusion cannot perceive.

Finally, the effect of different fusion levels on the same system was not explored in the reviewed articles. However, in certain scenarios this could highlight interactions between neural response patterns and modalities that could not otherwise be seen. Specifically, the difference between an hBCI relying on decision-level fusion and another, identical one, that relies on feature-level fusion, could be a factor in achieving better performance.

2.2.6 Performance metrics

Table 2.6 shows the metrics used in the selected articles. The most frequently used performance metrics included classification accuracy and other metrics derived from the confusion matrix, such as the true positive (TP) rate (also called sensitivity or recall), true negative (TN) rate (specificity), false positive (FP) rate (fall-out), TP/FP and positive predictive value (TP/(TP+FP)). Seven articles used the Information Transfer Rate (ITR) or the Practical Bit Rate (PBR).

Many articles used task-specific metrics, such as task completion time and number of successful trials. These metrics are tailored for each BCI paradigm and/or application, and therefore do not allow comparisons between studies. The interested reader is referred to [123, 181] for a review and guidelines on how performance metrics should be used and reported in BCI studies.

2.2.7 Subjective evaluation

Table 2.7 presents the subjective evaluation procedures used in the selected articles. Five articles based their subjective evaluation on the NASA TLX test, which evaluates 6 points: mental demand, physical demand, temporal demand, performance, effort and frustration. Some application-specific questionnaires were used [21, 166], as well as general questionnaires as to which conditions were the most difficult or most annoying, and how well the user performed.

2.2.8 Comparison of hybrid and non-hybrid solutions

Table 2.8 presents the reported advantages and disadvantages of hBCIs as compared to standard non-hybrid BCIs in the reviewed articles. This section specifically focuses on articles that compared a hybrid paradigm to a non-hybrid one.

The most often reported advantages of using a hBCI included a higher classification accuracy, lower error rate, or higher area under the ROC curve than in an equivalent non-hybrid BCI. This was the case for studies where the hybridisation was specifically aimed at improving the robustness or accuracy of the system through input redundancy or brain switch mechanisms [168]. Similarly, lower false positive rates were reported in two articles with sequential or brain switch systems [12, 162]. Indeed, higher accuracies are expected when choosing input redundancy, brain switch or error correction mechanisms. With input redundancy, either different brain activities are elicited by the same stimulus or task, or different stimuli/tasks are mapped to the same commands. In both cases, complementary information can be found from the analysis of the two or more brain activities, increasing the chances of a successful recognition of user intent. This can also improve the performance of the BCI in noisy environments, as found in [57]. When brain switch mechanisms are used, the overall classification accuracy will improve due to the decrease in false positive rate provided by the switch. Finally, error correction strategies based on the detection of the ErrP can also improve the overall classification accuracy [145].

Higher information transfer rates were reported in a few studies, in which the different modalities or brain activities were always mapped to different commands or groups of commands. For example, error correction based on the ErrP was shown to improve the ITR of a P300 speller in both healthy and paralysed users [146]. Task-related metrics were also improved by hybridisation in [136, 159, 161]. Interestingly, a few articles mentioned the tendency of hybrid paradigms to improve the performance of

users with poor unimodal accuracy as compared to users with already high unimodal accuracy [17, 65, 144, 169].

Apart from metric-related advantages of hBCIs, articles also mentioned the ability to simultaneously operate different commands (for example, in wheelchair control applications, turning left/right while accelerating/decelerating), improved artefact detection (in the case of the reviewed articles, this always concerned the detection of eye movement artefacts using eye tracking or EOG) and the flexibility allowed by switching from one modality to the other based on the level of fatigue of the user [22]. Finally, questionnaire results in [126, 128] did not show an increase in the perceived difficulty or workload level for the hybrid condition, despite the usual higher complexity of hBCIs.

Various disadvantages of hBCIs were also reported. First, certain studies did not find improvements in classification accuracy or information transfer rate when using a hybrid paradigm. This can occur when one brain activity or modality performs much worse than the others, effectively bringing down the overall accuracy; for example, when the BCI channel is added to a traditional input modality such as a joystick [166], or when ERD detection is added to a SSVEP paradigm in [131, 133]. Second, according to certain studies, this phenomenon can also be attributed to dual-task interference [131, 133, 134, 156].

This, however, depends heavily on which mental activities are used together: for example, in [131, 133], adding SSVEP to ERD degraded the individual ERD detection accuracy, while adding ERD to SSVEP did not affect the individual SSVEP detection accuracy; in [138, 148, 166], adding a mental activity or modality did not affect the individual classification accuracies. Third, hybrid paradigms can lead to an increase in selection time: this can be caused by the adoption of a brain switch mechanism [12], by the addition of the BCI channel to a faster traditional user input such as an eye tracker [126], or by the use of modalities with an longer inherent response delay such as NIRS [17]. Finally, hybridisation often makes the BCI significantly more complex to use [135, 138], and can also, for certain hybridisation paradigms, decrease the usability [143] or increase the level of difficulty [133, 166] as shown in subjective evaluation results.

Overall, the majority of articles that compared hybrid and non-hybrid solutions found a significant increase in performance metrics: 23 found that metrics such as classification accuracy and ITR improved, while eight found that these metrics stayed the same or decreased. While the objective measures showed hBCIs usually performing better, subjective measures (i.e., as collected in questionnaires) showed usability is not necessarily improving when using hBCIs. It is thus imperative to validate the hybridisation paradigms. Indeed, as explored in [143], all hybridisation paradigms are not equally useful, and some configurations might even produce poorer results than a non-hybrid BCI. Lorenz *et al.* (2014) compared ERP-ERP, ERP-MI and MI-ERP BCIs in a sequential paradigm in which a prosthesis was controlled by first choosing one of six possible movements, and then confirming or cancelling the selection using a binary selection. The authors found that the ERP-ERP and ERP-MI protocols were similar in accuracy and usability, but that the MI-ERP solution produced much worse results. This highlights the fact that some mental activities or modalities may not be well suited in certain contexts, such as MI for multiple

selection when using a binary protocol (hand vs. foot motor imagery). This also shows that hBCIs do not necessarily perform better than non-hybrid BCIs. The pros and cons of each hybrid paradigm must thus be well studied before making a choice. Nonetheless, effective improvements in system performance can be achieved with hBCIs.

2.2.9 Recommendations from the literature review

Based on the many study characteristics discussed above, there are a few important points that should be considered in future hBCI studies. First, it was observed that many studies applied a hybrid paradigm without objectively justifying its use, for instance by measuring how it compares to non-hybrid solutions. Therefore, whenever possible, hBCI performance and usability should be compared to that of a system where only a single modality or neural response pattern is used (for example when hybridisation allows redundant mapping aimed at increasing performance on a single task). This would provide a rational justification for the use of a hybrid paradigm and allow competing hybrid paradigms to be compared. Moreover, different configurations should be evaluated and compared for new hBCI paradigms in order to identify the optimal configuration. Indeed, as shown in [143], switching the neural response pattern-command mapping in a sequential paradigm alone can result in drastic changes in performance and usability.

Passive hBCIs, especially affective hBCIs, were poorly represented in the selected articles. However, considering the extensive amount of work done in affect recognition using physiological signals such as SCR and heart rate [182], multimodal hBCIs with such modalities are already one step closer to integrating affect recognition in their system. Moreover, as was highlighted in [3], the potential of user-state monitoring interfaces is greater than interfaces that provide device control, thanks to the former not competing against traditional user hardware such as a mouse or keyboard. The incorporation of affective dimensions in hBCIs, such as in passive or composite active/reactive-passive hBCIs, is thus a very relevant topic.

Next, certain aspects of the experimental protocol should be given more attention. Indeed, reporting, as well as controlling for the experience of BCI users, is essential to devise robust and stable systems that can be adopted by as many people as possible. Similarly, long-term studies such as [53] must be carried out to understand how BCI performance evolves with time given different paradigms.

As covered in Section 2.1.3, artefact handling procedures are useful to increase the signal-to-noise ratio of physiological data affected by artefacts, and help in achieving robust classification performance. Although typically time-consuming to implement, some implementations are validated, openly available and straightforward to use [183]. Upcoming hBCI studies should thus consider including such a step in their pipeline. Furthermore, two artefact handling approaches that were barely or not at all reported in the selected studies, but that would be of particular relevance for hBCI studies are 1) the use of multiple modalities to better detect (and correct) artefacts, such as eye trackers and accelerometers, and 2) rather

than discarding non-neurophysiological information from EEG and NIRS, using it to emulate a second modality such as EMG or plethysmogram. In this way, not only hBCIs can benefit from an improved robustness or increased number of possible commands, but this can also enhance their signal-to-noise ratio in one case, and provide additional information without increasing the complexity or cost of the necessary hardware in the other. We encourage future hBCI studies to consider these possibilities when designing their processing pipeline.

Many EEG and BCI datasets are currently available online, for example the ones released for the different BCI challenges in the last few years [184–186]. These datasets are useful for developing new signal processing and classification procedures, while also serving as benchmark datasets for the BCI community. Processing pipelines can thus be objectively compared and orient future research directions. However, similar datasets on multimodal data or multiple neural response patterns are not as easily available. A few multimodal datasets are available for free on the internet, but they are not specific to active/reactive brain-computer interfacing, and rather focus on affective dimensions in EEG and physiological signals [187, 188]. To allow for a concerted effort toward well validated systems, authors of upcoming hBCI studies should thus consider releasing their data to the rest of the community.

2.3 Mental task classification using unimodal and multimodal BCIs

Various mental tasks that recruit different parts of the brain, such as mental subtraction and mental rotation, were investigated in recent studies using EEG [30, 36, 45–53], NIRS [54–62], TCD [63, 64], NIRS-TCD [65], and fMRI [66]. The most frequently used mental tasks in these articles were mental subtraction, mental object rotation, various verbal fluency tasks, motor imagery and auditory imagery. These studies attempted either binary or multi-class classification of mental tasks, usually between tasks or against a resting state, and in most cases aimed at finding the combination of tasks that would yield the highest performance in a BCI context as measured by a classification metric.

Unimodal studies: EEG or NIRS

In the studies that evaluated four or more different tasks in EEG, combinations of a brain-teaser, that is a task involving a problem or puzzle that induces some degree of mental work, and a dynamic imagery task were always found to yield the highest performance. In [47], binary classification of mental singing and mental calculation (either addition or subtraction) was found to be in the top five of the best combinations in four of five subjects, with values above 93% in accuracy. Although aiming to differentiate a mental task from resting state instead of from another mental task, Faradji *et al.* found that a mental multiplication task could produce a maximal true positive rate above 70% while maintaining a zero false negative rate [48]. Research by Friedrich and colleagues has supported the combination of a brain teaser and a dynamic imagery task as an optimal pair numerous times [36, 51, 53]. In [36], pairs formed from a total

of seven different mental tasks were evaluated on nine subjects. Word association, mental subtraction, mental rotation and motor imagery were identified as the most discriminative tasks, leading to Cohen's κ (an inter-rater agreement metric) values above 0.7. The ERD/ERS patterns evoked by brain teasers and dynamic imagery tasks were found to exhibit different characteristics, which could explain why this type of combination is optimal [36]. In addition, the authors characterized the subjective appreciation of each task, finding that while ratings were highly variable and no significant differences between tasks were found, word generation received the best rating and mental subtraction received the worst. Similarly, in [51], a combination of mental subtraction and motor imagery was found to produce consistent inter-session performance in seven out of nine subjects (κ higher than 0.6). Finally, in [53], eight subjects were trained to control a 4-class BCI in which mental tasks were selected based on an individual basis. The most frequent task combinations included motor imagery, mental rotation and additional brain teaser and dynamic imagery tasks, leading to average performances varying between 44 and 84% accuracy.

Only two studies were found to look at classifying more than two tasks at once using NIRS. In [59], using the hemodynamic information from the prefrontal cortex it was shown that binary classification accuracies around 60% could be obtained with pairs formed of mental subtraction, word generation and mental rotation on 10 subjects. A more complete assessment of mental tasks using NIRS was reported in [6], using a full head coverage. Classification accuracies around 71% were found for combinations of motor imagery, mental multiplication and mental rotation tasks on seven subjects.

Multimodal studies: EEG-NIRS, NIRS-Peripheral signals and NIRS-fTCD

Some studies explored the effect of adding supplementary modalities to the classification of various mental tasks to either improve the number of classifiable tasks or improve the robustness of a precise task. Combinations of NIRS and EEG for recognizing motor imagery or execution tasks have been studied following three different approaches. Fazli *et al.* used a classifier fusion procedure based on the individual classification of EEG, HbO and HbR data to distinguish left from right hand motor execution and imagery [17]. Using this approach, an average 5% increase in accuracy was obtained for 13 out of 14 subjects for the motor imagery task when EEG and HbO were used simultaneously, yielding an average accuracy of 83.2%. The authors recognized the drawbacks of the long hemodynamic response delay that typically precludes the use of NIRS in practical BCIs. A different approach was explored by Khan *et al.* to increase the number of input commands in the context of four-direction movement control [18]. While the left and right movements were controlled with motor imagery as measured with EEG, the forward and backward movements were controlled with either counting or mental subtraction tasks measured in NIRS. A binary classification of each task against rest was used. Average classification accuracies of 94.7% and above 80% were obtained for EEG tasks and NIRS tasks, respectively. Finally, in [19], NIRS was used to detect the occurrence of a motor imagery task and trigger its classification as left or right imagery using EEG. This approach led to a low False Positive rate of 7% and a True Positive rate of 88%.

NIRS has also been combined with peripheral physiological signals such as heart rate, respiration, blood pressure, skin temperature and SCR measurements. In [57], the impact of environment noise on music imagery detection was assessed using NIRS, physiological signals, and a combination of the two. An average accuracy of 83% over eight participants was obtained when NIRS was used in conjunction with physiological signals, corresponding to a 12% gain from a NIRS-only system. Using a similar methodology, Zimmermann *et al.* explored the detection of a motor execution task [169]. The authors found that adding physiological signals to NIRS significantly improved classification accuracies by around 9% average in seven subjects. Joining the hemodynamic information obtained from NIRS with the one obtained from fTCD, a modality that measures blood flow velocity in the cerebral main arteries, was also shown to be beneficial in the classification of a mental task. Indeed, in [65], this approach helped improve unimodal classification of a verbal fluency task by around 7% average in nine subjects.

2.4 Proposed improvements

In light of this introduction, the purpose of this thesis is to gain further ground in the investigation of new BCI control tasks, by adopting a hybrid BCI approach in which NIRS and EEG are simultaneously recorded. We seek to complement previous studies such as [36, 61] by carefully analyzing the functioning of such a hybrid system and assessing whether it can provide a gain in performance over standard unimodal approaches. Windowing parameters, as well as the dynamic evolution of classification performance, are also explored to further understand how hBCIs can be improved. Finally, an exploratory analysis is carried out to identify neurophysiological features that could be used to predict the subjective experience of a BCI user, potentially leading to user-aware passive BCIs.

Table 2.4: Extracted features in the selected hBCI studies.

Modality	Neural response pattern	Feature	Articles
EEG	ERP (P300/ErrP)	Single-trial amplitude values	[137, 139, 141–143, 145, 146, 150, 153, 154, 164]
		Averaged amplitude values	[138, 140, 145, 148, 151, 152, 156, 158]
		Spatial-filtered amplitude values	[155]
		Mean spectral power across epochs + Inter-trial phase coherence	[149]
		Frequency features from the Synchrosqueezing Transform (SST)	[147]
	SSEP (SSVEP/SSSEP)	(Logarithmic) band powers (with or without harmonics)	[21, 128, 131, 134, 148, 152, 168]
		Amplitude values	[130, 132]
		Harmonic Phase Coupling	[131]
		Harmonic Sum Decision	[12, 131]
		Mean spectral power across epochs + Inter-trial phase coherence	[149]
		Power spectral density of CCA-filtered channels	[150]
		Correlation components from CCA	[135, 136, 154, 155]
		Power spectral density of the Minimum Energy Combination for fundamental and first harmonic	[151]
	MI/ME	(Logarithmic) band powers	[12, 19, 22, 128–133, 138, 159, 161–163, 166, 167]
		Logarithmic variance of CSP components	[19, 51, 135, 137, 139, 140, 144, 158]
		CSP features; meta-channels obtained with SpecCSP	[17, 36, 126, 134, 134, 136, 141, 141–143, 155, 157]
		Average peak amplitude	[18]
		EEG-EMG coherence values	[160]
	Alpha waves	Band power over alpha frequencies	[21]
	Mental tasks	Logarithm variance of CSP components	[36, 51, 53]
		Band powers (from CSP components)	[165]
	Passive	Cross-correlation patterns (Pearson's coefficients) between the different channels	[170]
NIRS	MI/ME	Averaged delta concentration	[17, 19]
		Coefficients from linear combination of NIRS channels	[169]
	Mental tasks	Position-averaged DC values	[57]
		Averaged delta concentration	[18]
		Slope of Hb, HbO and/or HbT	[65, 168]
Eye tracker	-	Pupil information	[164]
EMG	-	Averaging of amplitude values	[129]
		Hilbert transform components	[159, 161]
		EEG-EMG coherence values	[160]
EOG	-	Amplitude values of a bipolar channel	[157]
		Correlation components from CCA	[158]
fTCD	-	Slope, mean, change from initial to end of trial and range of amplitude values	[65]
Other physiological signals	-	Concatenated amplitude values	[57, 169]

Table 2.5: Fusion level of the selected hBCI designs. Systems with more than one function sometimes have different fusion levels for each function [135, 139, 141, 142, 155, 158].

Fusion level (# of articles)	Articles
Data-level (7)	[36, 51, 53, 144, 147, 149, 160]
Feature-level (12)	[57, 65, 128, 130, 131, 133, 139, 141, 142, 155, 169, 170]
Decision-level (27)	[17, 19, 129, 134, 135, 137, 139–143, 145, 146, 148, 150–154, 156–159, 161, 164, 165, 168]
Semantic-level (15)	[12, 18, 21, 22, 126, 132, 135, 136, 138, 155, 158, 162, 163, 166, 167]

Table 2.6: Reported performance metrics in the selected articles. Studies may appear in more than one category.

Performance metrics (# of articles)	Articles
Classification accuracy (30)	[17, 18, 22, 53, 57, 65, 126, 128–133, 135, 136, 138, 140, 141, 143–145, 147, 150, 152, 153, 155–157, 160, 168]
ROC curve, Area Under the Curve (3)	[21, 164, 165]
Other metrics derived from the confusion matrix (13)	[12, 19, 57, 65, 129, 143, 145, 151, 158, 161–163, 169]
Bit rate (14)	[19, 133, 135, 146, 148–151, 153–157, 168]
Cohen's Kappa Coefficient (3)	[36, 51, 162]
Task-specific metric (26)	[126, 132, 135, 137, 139, 157, 158, 165] [19, 151, 160, 161] [167] [22, 132, 134, 137, 139, 140, 142, 143, 149, 153, 155, 159, 166]
None (1)	[170]

Table 2.7: Type of subjective evaluation of the hBCI designs performed by the users.

Subjective evaluation (# of articles)	Articles
NASA TLX (5)	[126, 139, 142, 143, 167]
Game experience questionnaire (1)	[21]
Slater-Usoh-Steed presence questionnaire for Virtual Environments (1)	[166]
Questionnaire on difficulty, usability, performance, etc. (8)	[36, 53, 128, 131–134, 143]
None (41)	[12, 17–19, 22, 51, 57, 65, 129, 130, 135–138, 140, 141, 144–165, 168–170]

Table 2.8: Advantages and disadvantages of hBCIs as compared to non-hybrid BCIs, as reported in the selected articles.

	Description (# of articles)	Articles
Advantages of hBCIs	Higher classification accuracy, lower error rate, or higher area under the ROC curve (15)	[17, 57, 65, 128, 129, 131, 144, 145, 149–152, 164, 168, 169]
	Higher information transfer rate (8)	[135, 146, 149, 150, 153, 154, 156, 168]
	Lower false positive rate (2)	[12, 162]
	Higher task-related metric score (3)	[136, 159, 161]
	Allow simultaneous commands (3)	[134–136]
	Improve artefact detection (3)	[156, 162, 163]
	Allow switching between modalities to limit user's fatigue (1)	[22]
	No increase in perceived difficulty level or workload (2)	[126, 128]
	Less frustrating/more fun (2)	[126, 166]
Disadvantages of hBCIs	Not significantly different/lower classification accuracy or information transfer rate (5)	[131, 133, 153, 159, 166]
	Dual-task interference (4)	[131, 133, 134, 156]
	Increased selection time (and thus decreased TPR and FPR) (3)	[12, 17, 126]
	More complex/difficult to control (4)	[133, 135, 138, 166]
	Poorer usability pragmatic qualities (1)	[143]

Chapter 3

Methods and Materials

3.1 Participants

Twelve participants (five females, three left-handed, 20-31 years old, mean 24.6 year old), fluent in English and/or French, took part in our NIRS-EEG study. Participants had to complete three separate recording sessions of two to three hours, inside a period of three to five weeks. Participants declared having no history of neurological disorders and had no previous experience with BCIs. A monetary compensation was given after each completed session. Each participant agreed with the terms and conditions of the study, which was approved by the university ethics committee. The data from two participants was rejected, as it was highly artefacted; data from a third participant was also rejected because they did not complete the three required sessions. Therefore, the data from a total of nine participants was used in this study.

3.2 Mental tasks

Participants were asked to perform seven different types of mental task that are believed to elicit distinct neural response patterns, based on previous studies using EEG and NIRS [36, 61].

Mental rotation (ROT) Two 3-dimensional L-shaped figures, either identical or mirrored, but in each case in a different state of rotation, were presented to the participants. Participants had to imagine rotating one of the two figures in order to establish whether they were the same or if they were mirrored images.

Word generation (WORD) Participants had to generate as many words as possible starting with a randomly chosen letter presented on the screen. Words in either English or French, depending on the participant's chosen language, were requested.

Mental subtraction (SUB) Participants had to perform successive subtractions of two one- or two-digit numbers from a three-digit number (e.g., 214 - 9 = 205, 205 - 13 = 192, etc.).

Mental singing (SING) Participants had to imagine singing a song that they had chosen beforehand, if possible with lyrics, while focusing on the emotional response it elicited.

Mental navigation (NAV) Participants had to imagine walking from one point to another in their current or a previous home, while focusing on their spatial orientation (e.g., walking from their bedroom to the refrigerator).

Motor imagery (MI) Participants had to imagine performing a finger tapping task with their right hand.

Face imagery (FACE) Participants had to imagine the face of a friend, as recalled from a picture they were asked to bring to the recording session and memorize.

Following Friedrich *et al.*'s description of task types [36], we classify word generation and mental subtraction as brain-teasers, since they require problem solving skills; mental singing, mental navigation and motor imagery as dynamic imagery tasks; and face imagery as a static imagery task. We also classify mental rotation as a brain-teaser instead of a dynamic imagery task because in our case the subjects had to produce an answer (i.e., whether the figures matched or not). This modification was motivated by the mental rotation literature, which often uses this version of the task [67, 68].

Before the recording started, participants were first guided through each task and asked to complete them in an overt manner. This step was used to make sure each participant performed the tasks appropriately and in a similar manner. For example, for the mental rotation task, participants had to say aloud whether the two figures were the "same" or "mirrored", for the mental subtraction task, they had to give the experimenter their final answer, etc. Participants were then asked to repeat the same tasks but in a covert manner to prepare them for the way they would have to perform them during the experiment.

3.3 Experimental paradigm

The experimental paradigm for a single trial of our study is summarized in Figure 3.1. A complete session consisted of four subsessions in which each mental task type was randomly repeated four times, yielding a total of 28 task completions per subsession. Participants also had to complete a subjective evaluation questionnaire between the second and third subsession of each session. Each subsession started and finished with a 30-s baseline period in which participants were asked to remain in a neutral mental state and fixate the cross at the center of the screen. Before each trial, a 3-s countdown screen identified the task to be performed next using the associated pictogram as shown in Figure 3.1.

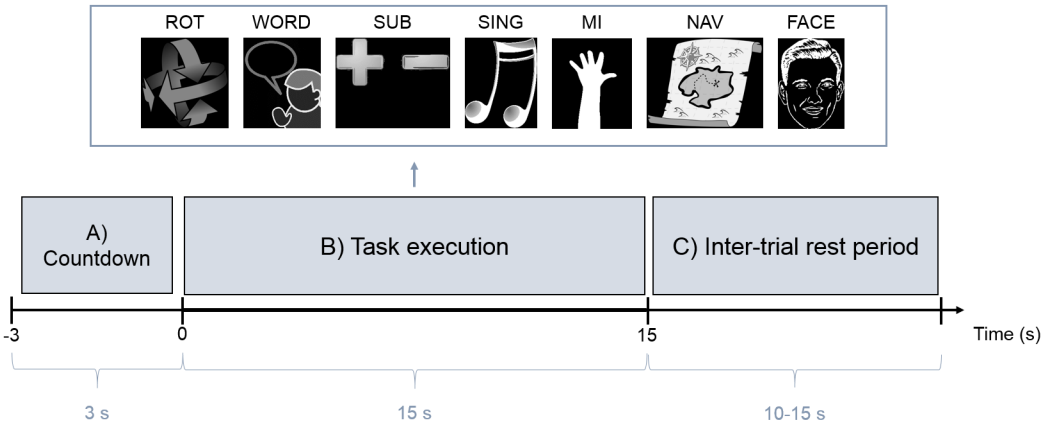

Figure 3.1: Diagram of a trial in the experimental paradigm. A trial is composed of A) a 3-s countdown period in which participants are instructed about the upcoming mental task, B) an imagery period where they execute the given task for 15 s, and C) a randomized 10- to 15-s rest period.

Once the countdown was over, participants had to execute the required mental task for a period of 15 seconds. Instructions were given to carry out the tasks as many times as possible, and to start tasks such as rotation, subtraction and navigation over again if completed before the end of the 15-s period. Each trial was then followed by a rest period of random duration, sampled from a uniform distribution between 10 and 15 s, in which participants were asked to continue minimizing movements, but were allowed to blink, swallow, etc. if need be. The duration of the rest period was randomized to ensure that participants could not expect the exact start of the next trial, as well as to avoid synchronization of systemic processes in NIRS with the paradigm. Once a subsession was over, participants were allowed to take as much time as desired to stretch a bit, drink or eat a snack before resuming the experiment. The stimuli and questionnaire were both implemented using the Presentation software (Neural Behavioral Systems, USA).

3.4 Data collection

3.4.1 EEG and physiological signals

EEG data was recorded using an ActiveTwo system (Biosemi B.V., Amsterdam, The Netherlands) with 62 probes (plus two mastoids, and CMS and DRL electrodes) and four EOG electrodes. No online filtering was applied. A standard 10-10 system was used for electrode placement, but without AF7 and AF8, whose holders were used for NIRS probes instead (see Figure 3.2).

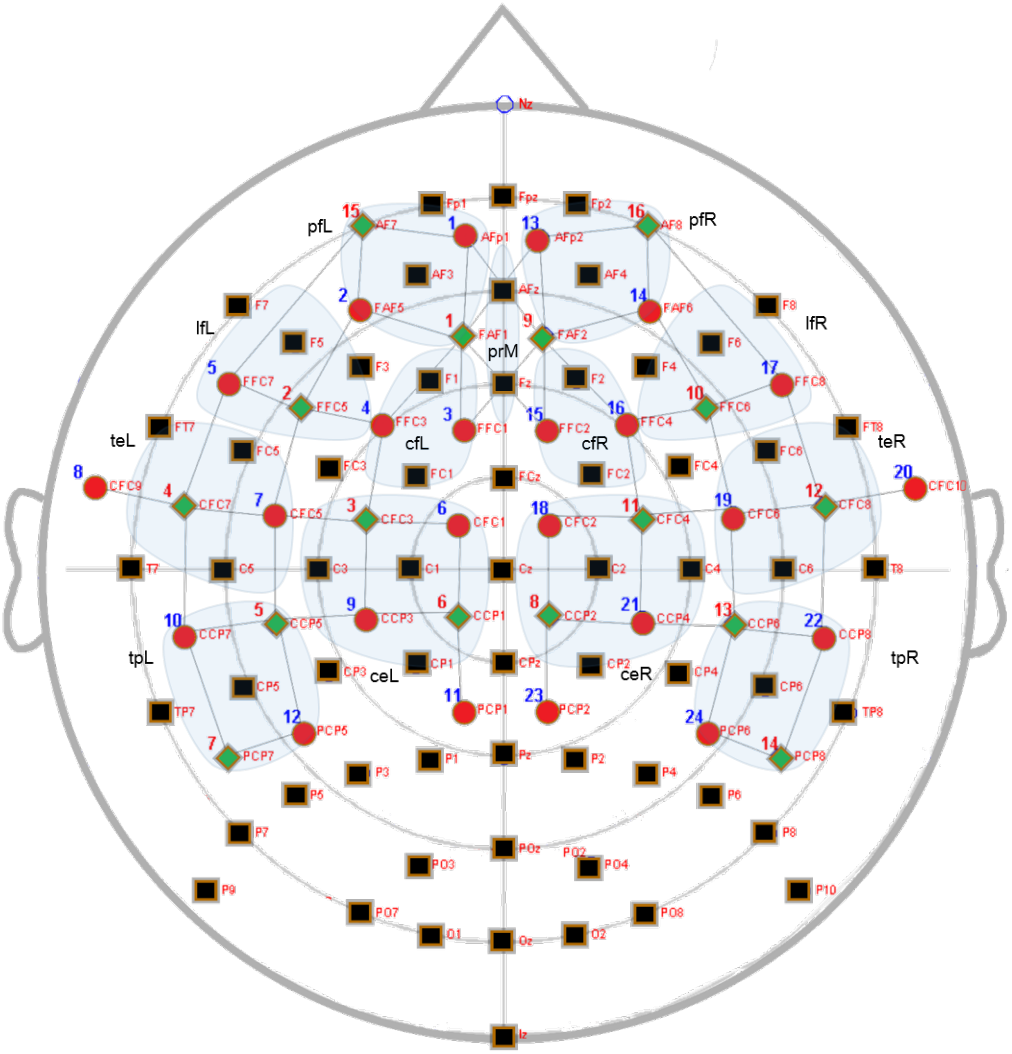

Figure 3.2: EEG and NIRS topology used in this study (adapted from [189]). EEG electrodes (black rectangles), NIRS detectors (red circles) and NIRS sources (green diamonds) were placed following the 10-10 system. NIRS channels are represented by dark straight lines connecting the sources to the detectors. Brain regions used to compute artificial NIRS channels are grouped in light blue.

Peripheral signals were simultaneously recorded using the same amplifier: respiration (belt around the chest), SCR (two electrodes on the middle and ring fingers), skin temperature (on the little finger) and plethysmograph (on the index finger). Due to technical malfunctioning, peripheral signals were unfortunately not recorded during half of the sessions and thus were not used for this analysis. All signals were digitized at a 512 Hz sampling rate.

Table 3.1: Questionnaire items and description based on the NASA TLX test. The first five questions were given a rating between 1 and 10.

Dimension	Question
Mental demand	How mentally demanding was the task?
Temporal demand	How hurried or rushed was the pace of the task?
Performance	How successful were you in accomplishing what you were asked to do?
Effort	How hard did you have to work to accomplish your level of performance?
Frustration	How insecure, discourage, irritated, stressed and annoyed were you?
Task ranking	Which are your preferred tasks, in order of importance?

3.4.2 NIRS

NIRS data was recorded using a NIRScout system (NIRx Medical Technologies, Los Angeles, USA), with 16 sources (wavelengths of 760 and 850 nm) and 24 detectors. Optodes were placed together with EEG electrodes on the same cap, as shown in Figure 3.2. The frontal, central, temporal and parietal regions were targeted by the montage used, following the extended 10-5 system. Coverage was not extended to the occipital lobe due to the low quality of NIRS signals in this region and because of the restricted number of available optodes. Source-detector pairs separated by approximately 3 cm were used as channels, giving a total of 60 NIRS channels, each sampled at 4.46 Hz. Amplifier gains were adjusted following an automatic calibration procedure handled by the NIRStar recording software.

3.4.3 Questionnaire and subjective ratings

At each recording session, participants were asked to fill out a questionnaire reporting their appreciation of the tasks. The questions were based on the first part of the NASA TLX test [69], with the French version by [70]. Additionally, participants were asked to rank the mental tasks in order of preference. Table 3.1 shows the various items that were measured.

3.5 EEG analysis

3.5.1 Preprocessing

The raw EEG data was preprocessed using EEGLAB [71], a freely available MATLAB toolbox. The data was imported and referenced to Cz before being downsampled to 256 Hz. The referenced data was bandpass-filtered between 0.5 and 100 Hz and notch-filtered at 60 Hz to reduce power line interference. Bad channels were visually assessed and removed across trials of a same session. Epochs of 35 seconds with 15 and 10 seconds of baseline before and after task execution respectively were extracted for each trial, as well as for rest periods at the beginning and end of each subsession. Epochs contaminated by

large movements and physiological artefacts were visually identified and rejected. Further signal cleaning was performed by computing the ICA decomposition of the epoched data, and then applying a semi-automatic method to detect and reject eye movement artefact components and other noisy components [72]. Bad channels were re-interpolated and all channels were re-referenced to the common average of all electrodes. Finally, baseline correction was applied to each epoch using the 300 ms period before the beginning of the task.

3.5.2 Descriptive analysis

ERD is a phenomenon occurring when idle parts of the brain become active following some event or stimulus. A specific rhythmical activity can then be measured, such as μ and β waves over the primary motor cortex in motor execution and imagery tasks [190]. Similarly, ERS occurs when this activity ceases and the recruited brain regions return to an idle state. The inter-trial variance method proposed in [73] was used to compute the ERD/ERS values for each type of mental task and baseline over all three sessions for each participant.

3.5.3 Feature extraction

Previous EEG studies mainly used band powers and Common Spatial Patterns-based (CSP) features to describe the neural activity patterns induced by mental tasks such as motor imagery (see Section 2.2.4). In our case, since we are especially interested in the interpretability of the extracted features, we focused our classification analysis on classical power bands. Additional oscillatory process-related features that were shown to be useful in other EEG studies were also computed, but only used in the analysis of correlation with subjective ratings [74, 75]. Therefore, four different types of features were extracted for each electrode, yielding 1894 features in total:

- Log-power in frequency bands and ratios between them
- Amplitude modulation in frequency bands
- Amplitude and phase coherence over inter-hemispheric electrode pairs
- Global phase synchrony

The nomenclature used to describe these features is presented in Appendix B. Each mental task trial was subdivided into non-overlapping time windows of either 1, 2, 3, 5 or 15 seconds, or overlapping windows of 7.5 or 10 seconds, used to extract the above-mentioned features (see Table 3.2).

Table 3.2: Window sizes for extraction of EEG and NIRS features.

Window duration (s)	Number of windows	Overlap (s)
1	25	0
2	12	0
3	7	0
5	5	0
7.5	4	3.25
10	3	5
15	2	10

Log-power in frequency bands and ratios between them

First, log-power features were extracted in the following 10 frequency bands: δ (0-4 Hz), θ (4-8 Hz), α_{low} (8-10 Hz), α_{high} (10-12 Hz), β_{low} (12-21 Hz), β_{high} (21-30 Hz), γ_{low} (30-60 Hz), γ_{high} (60-80 Hz), θ to β (4-30 Hz), and total spectrum (0.1-100 Hz). To do so, a fast Fourier transform was applied on the extracted windows. The absolute spectrum was then squared and summed in each frequency band defined above. Moreover, the following ratios of band powers were computed: $\frac{\alpha_{total}}{\beta_{total}}$, $\frac{\theta}{\beta_{total}}$, and $\frac{\gamma_{high}}{\beta_{total}}$. This yielded a total of 806 features per window, which were finally log-transformed.

Amplitude modulation rate-of-change

Amplitude modulation features encode information about the interaction of different rhythmic activities in the brain [74], and were shown to be more useful than standard spectral peak parameters in distinguishing healthy individuals from patients suffering from Alzheimer's disease [74, 75]. The relatively new utilization of these types of features for EEG data means that very few applications have been tested yet. We hypothesize that the hierarchical spectral information provided by these features can be useful to describe other cognitive processes such as subjective experience.

Amplitude modulation features were computed following the methodology proposed in [74]. First, EEG signals were band-passed in the following bands: δ, θ, α and β. Second, the envelope of each of these four band-passed signals was obtained using the Hilbert transform. Finally, the new envelope signals were further filtered in the same previously mentioned frequency bands (δ, θ, α and β). The average power of these new amplitude-modulated signals over the extracted windows is computed and used as a feature. This procedure yielded nine different features per electrode (written as <frequency band> modulated by <other frequency band>): δ m-δ, θ m-δ, θ m-θ, α m-δ, α m-θ, β m-δ, β m-θ, β m-α, and β m-β. This yielded a total of 558 features per window.

Amplitude and phase coherence

The magnitude-squared coherence (MSC) and the phase coherence (ϕ_{coh}) measure the similarity between the power spectrum of two signals. These features were previously used for studying language [76], working memory [77] and neurological diseases [74], amongst others. The magnitude coherence for a pair of signals $x(t)$ and $y(t)$ is computed using the spectra $X(f)$ and $Y(f)$ of each signal:

$$MSC(f) = \frac{|X(f)Y^*(f)|^2}{|X(f)||Y(f)|}, \tag{3.1}$$

where the complex conjugate of $Y(f)$ is written as $Y^*(f)$. The square-root of the numerator is called the complex coherence, and is also used to compute the phase coherence:

$$\phi_{coh}(f) = arg(X(f)Y^*(f)), \tag{3.2}$$

where the function arg returns the imaginary part of the complex coherence.

We extracted the MSC and the ϕ_{coh} over 26 inter-hemispheric electrode pairs (Fp1-Fp2, AF3-AF4, F7-F8, F5-F6, F3-F4, F1-F2, FT7-FT8, FC5-FC6, FC3-FC4, FC1-FC2, T7-T8, C5-C6, C3-C4, C1-C2, TP7-TP8, CP5-CP6, CP3-CP4, CP1-CP2, P9-P10, P7-P8, P5-P6, P3-P4, P1-P2, PO7-PO8, PO3-PO4, O1-O2) for each frequency band previously defined (Section 3.5.3), yielding a total of 520 features per window.

Global Phase Synchrony (GFS)

The phase synchrony between a certain number of electrodes in specific frequency bands can be used to evaluate the global synchronization of brain regions [78]. Although GFS has mostly been used to study schizophrenia and Alzheimer's disease [75, 78], we investigate the use of this non-classical feature for capturing different neuronal oscillatory processes. Details on the computation of GFS can be found in [78]. One feature was computed in each band defined in Section 3.5.3, yielding a total 10 additional features per window.

3.6 NIRS analysis

3.6.1 Preprocessing

NIRS preprocessing was performed using the open source toolbox HOMER2 [79]. First, raw light intensities were converted to OD by computing the negative logarithm of the normalized intensities (using the average value of each channel over the entire recording). Second, channels with a low signal-to-

noise ratio were identified by correlating the bandpass filtered OD (between 0.8 and 1.2 Hz) of channels S15-D1 and S15-D2 on the left and channels S16-D13 and S16-D14 on the right, with each left side or right side channel, respectively. These four frontal channels being the shortest in the montage used, and usually being clear of any hair, they are expected to carry a clear cardiac pulse in the 45 to 100 BPM range, which corresponds to standard resting heart rate frequencies. Channels which significantly correlate with these short-distance channels in this frequency range (p-value below 0.05) are thus expected to be of good quality, since they carry physiological information. Channels were thus rejected if the median of the Pearson's correlation p-value across all epochs of one session was higher than 0.05. Third, the remaining channels' OD were bandpass filtered between 0.01 and 0.30 Hz, and converted to concentration changes in oxygenated, deoxygenated and total hemoglobin (HbO, HbR and HbT) using the modified Beer-Lambert law with a partial pathlength factor of 6.

Finally, 13 artificial channels were computed by averaging the amplitude of neighbouring channels in the latero-frontal, centro-frontal, temporal, temporo-parietal, central and prefrontal regions. These steps produced 73 channels, each one measuring HbO, HbR and HbT, giving a total of 219 measurement channels for NIRS.

3.6.2 Descriptive analysis

The average HbO, HbR and HbT responses over sessions and participants were computed and plotted on topographical maps to reveal spatial patterns of hemodynamic activation for each mental task.

3.6.3 Feature extraction

Classification of NIRS data is often based on simple features, such as Δ[HbO], Δ[HbR] and Δ[HbT] averaged over time, their slope, or even the averaged raw light intensity values [14]. In this study, the following features were extracted using the same window sizes as for EEG feature extraction on each NIRS channel (HbO, HbR and HbT, including artificial channels): average amplitude, latency (in seconds) for maximum and minimum amplitude in the window (peak and valley), amplitude difference between average of the first half and average of the second half of the window, and coefficients of a linear, quadratic and cubic fit of the window. Only the average amplitude features were used in the classification tasks; the other features were used in the correlation analysis only. The nomenclature used to describe these features is presented in Appendix B.

Whereas the average amplitude feature only gives static information about the hemoglobin levels, the four other types of features give more dynamic information, and can be useful to discriminate hemoglobin responses that differ in temporal evolution. On a per subject basis, NIRS channels that were rejected for the three recording sessions during preprocessing were not used to compute features, while the missing

values in partially rejected channels (for example, rejected in only one of three sessions) were filled with the median of the other sessions. This yielded a total of 2849 NIRS features.

3.7 Classification of EEG, NIRS, and EEG-NIRS

Feature subsets taken from the extracted EEG and NIRS datasets were then used to train a binary classification model over pairs of tasks. For EEG, band power features (θ (4-8 Hz), low α (8-10 Hz), high α (10-12 Hz), low β (12-21 Hz), high β (21-30 Hz), θ to β (4-30 Hz), and total (0.1-100 Hz)) were selected (558 features), whereas for NIRS, average amplitude values for Δ[HbO], Δ[HbR] and Δ[HbT] were used (219 features). Pre-selection of a subset of the features was done to mitigate over-fitting problems caused by the curse of dimensionality. First, the training procedure was applied to the EEG and NIRS reduced datasets separately to assess the individual performance of each modality. Then, the same procedure was applied on a merged dataset combining features from both EEG and NIRS to assess the impact on classification performance.

To further avoid the over-fitting problems linked to high-dimensional datasets, especially in cases where only a few training cases are available, feature selection procedures are required to select the most informative features prior to or during classifier training. In this study, we used a linear kernel Support Vector Machine (SVM) classifier combined with a sparsity-inducing l_1 penalty term to control the number of features used in the model. As opposed to filter feature selection methods, this embedded procedure allows the efficient discovery of interdependent features, while also making better use of the available data by avoiding an additional partitioning [80]. The rotational invariance of the l_1 penalty also means that the number of samples needed to correctly learn a model grows logarithmically with the number of irrelevant features, instead of linearly as is the case for models using l_2 regularization [81]. This results in the l_1 penalty being more robust than l_2 when given a high number of irrelevant features [81]. Combined with the robustness of SVMs in high dimensional spaces [191], our embedded feature selection-classification procedure is designed to minimize the detrimental effects of our high dimensional dataset with few examples.

A default value of 2.0 was used for the hyperparameter C that controls the balance between the data-dependent loss and the l_1 penalization of the weights. To account for the varying number of data points in each classification task, C was divided by the number of examples in the training set, yielding a maximal value of 0.023 when all 96 epochs were conserved [117].

Tasks from the three recording sessions of a single participant were shuffled and divided into 10 partitions with an equal number of repetitions for each type of task. Since there will typically be a delay between the cue to execute a mental task and its actual execution, and since participants are likely to get tired and stop executing the task before the end of the 15-s epoch, each type of time window was analyzed independently, that is, a new classifier was trained for each time window. This makes it possible

to see how the classification changes as the task is executed. The classifier training was carried out using a 10-fold cross-validation procedure, in which nine partitions are used for training, and the remaining one is used for validation. Partitions were created so that an equal proportion of each class is found in each fold. For each division of training and validation sets, each feature was then individually Z-score normalized by subtracting its mean and dividing by its standard deviation as evaluated on the training set. The role of normalization in this case is two-fold: first, it prevents the learning algorithm from giving more importance to some features only because their magnitude is higher than others; second, it means that the absolute weights the SVM learns represent the importance of each feature (see paragraph below on feature importance). This normalization procedure was repeated for each combination of the 10 folds to avoid leakage of the validation set into the training set. The average validation score was then computed for each combination. Finally, the classifier was retrained on all 10 folds to obtain the final weights w.

Since our study is targeted towards the eventual conception of a BCI, in which it is important that time delays be minimized as much as possible to provide higher information transfer rates, we focus the bulk of our analysis on models trained with features extracted on windows of one second, as this allows a very short delay between the realization of the mental task and the output of the system. Moreover, using one-second windows allows a fine-grained analysis of the temporal evolution of selected features and of classification performance, which can uncover interesting physiological insights. However, we still provide an analysis of the effect of the feature extraction time window size on classification performance.

Performance evaluation

The performance of the classifiers was evaluated using Cohen's kappa inter-rater agreement coefficient (κ) as in [36], which measures the agreement between two raters who classify N examples into mutually exclusive categories. In our case, the two raters correspond to 1) the true label y_i attributed to an example of the dataset during data collection and 2) the output of the classifier $\hat{y}_i$. The relationship between the two raters can be expressed using a confusion matrix H shown in Table 3.3.

Table 3.3: Confusion matrix H for a binary classification experiment. Each entry $H_{i,j}$ in the table corresponds to the number of examples in the dataset that are classified as being from class 0 or class 1 ($\hat{y}$), depending on their true label y.

		$\hat{y}$	
		Class 0	Class 1
y	Class 0	$H_{1,1}$	$H_{1,2}$
	Class 1	$H_{2,1}$	$H_{2,2}$

Cohen's κ is useful in this context since the number of repetitions available for each type of task after bad epoch rejection is not equal, leading to an unbalanced number of examples in the dataset.

Cohen's κ is computed using the equation ([36]):

$$\kappa = \frac{p_0 - p_e}{1 - p_e}, \tag{3.3}$$

where p_0 is the classification accuracy and p_e is the probability of chance agreement. These are further defined as

$$p_0 = \frac{\sum_i H_{i,i}}{N} \tag{3.4}$$

and

$$p_e = \frac{\sum_i n_{:,i} \times n_{i,:}}{N^2}, \tag{3.5}$$

where the $H_{i,i}$ are the diagonal elements of the confusion matrix H, N is the total number of elements in H, and $n_{:,i}$ and $n_{i,:}$ are the sums of each column and each row of H, respectively. In the case of a perfectly balanced problem, kappa values of 0 and 1 are equivalent to 50% and 100% accuracy, respectively.

We focused our analysis on three ranges of κ values: 0.0-0.4 (low performance), 0.4-0.7 (satisfactory performance), and 0.7-1.0 (high performance). We define the low performance range as κ between 0 and 0.4, corresponding to 50% and 70% accuracy in a balanced setting, while the high performance threshold of 0.7 corresponds to an accuracy of 85%. An accuracy of 70% is usually considered to be the minimal performance a BCI should reach to be usable [15].

Feature importance

The best features can also be analyzed using the SVM weights of these models as a feature ranking metric [82]. Since the features are normalized before training, the weights of the linear model effectively represent the importance of each feature. A feature with a high absolute weight can thus be thought of as being more important than one with a weight closer to 0. In the present study, the absolute value of the weights of a particular feature were averaged and then ranked against the other features. Specifically, the spatial location of these features, as well as the relative importance of neuroelectric and neurohemodynamic feature types, were assessed in this work.

3.8 Exploratory analysis of the correlation between neurophysiological features and subjective ratings

We then explored the explicative power of each previously extracted EEG and NIRS feature in regard to the subjective dimensions measured with the NASA TLX questionnaire. This analysis was done on features extracted using 15 seconds of data, as they provide more information than shorter windows (especially for NIRS, considering its hemodynamic delay) and would be useful in a passive BCI context,

which typically does not require high ITRs. First, the average of a feature during a specific task was matched to the five ratings (mental load, temporal load, performance, effort and frustration) given to that task in that session. The process was repeated for each task, session, and subject. This yielded a table with 189 examples (7 tasks/session $\times$ 3 sessions/subject $\times$ 9 subjects) and 4751 dimensions (4746 features $+$ 5 subjective rating dimensions).

Next, Spearman's rank correlation (ρ_{spear}) was computed for each feature and subjective dimension pair. This correlation coefficient, as opposed to Pearson's linear correlation coefficient, does not assume any parametric relationship between the two variables, but only that a monotonically increasing function could perfectly explain their relationship. This is a more general case and might thus capture a significant effect where $\rho_{pearson}$ fails to see a relationship. Following a Bonferroni correction of the obtained correlation p-values to reduce the false discovery rate, the absolute correlations and the p-values were thresholded at 0.4 and 0.05, respectively, to identify features that were significantly correlated with the subjective ratings. Features that were significantly correlated with the subjective dimensions were then reported for each subjective dimension.

Chapter 4

Experimental Results

In this chapter, the results of the experiments described in the Chapter 3 are highlighted. More precisely, the following are detailed: the physiological data preprocessing outcomes, the classification results including the analysis of the classifiers obtained, the descriptive analysis of physiological data to support the classification findings, and the analysis of correlations between the extracted features and the subjective ratings.[1]

4.1 Preprocessing of physiological data

As a first step, the collected neurophysiological data was preprocessed and the data quality was visually assessed. The results of the preprocessing are summarized in Tables 4.1 and 4.2. The decision to reject the data from participants S02 and S08 was made based on the substantially lower number of valid trials, and their higher overall drowsiness as observed during the recordings. Moreover, data from participant S12 was rejected because only one session out of the three required was recorded. This left nine participants with at least 53 good EEG channels, 27 good NIRS channels, and 274 good trials (81.55% of the recorded trials). Missing EEG channels were re-interpolated using neighbouring electrodes, whereas missing NIRS channels were simply ignored. The data from these nine participants was used for the subsequent analyses.

4.2 Subject-wise mental task classification

In this section, the results of EEG-only and NIRS-only classifications are first described, followed by the results for the NIRS-EEG fusion classification. Specifically, the peak classification performance across

[1] These results constitute the material of a journal paper, in preparation, to be submitted to the IEEE Transactions on Neural Engineering and Rehabilitation [95].

Table 4.1: Number of rejected EEG and NIRS channels following the preprocessing procedure. A red cell indicates the participant was excluded from the analysis.

Participants	Number of rejected EEG channels			Number of rejected NIRS channels		
	session 1	session 2	session 3	session 1	session 2	session 3
S01	1	0	1	14	6	4
S02	1	0	3	24	18	19
S03	7	6	3	6	7	17
S04	1	8	6	25	33	23
S05	3	3	3	3	16	4
S06	6	0	6	0	0	3
S07	8	9	5	0	0	4
S08	5	6	4	8	7	3
S09	3	7	6	17	9	8
S10	2	7	8	0	1	0
S11	3	6	5	6	8	7
S12	3	-	-	4	-	-

Table 4.2: Number of valid trials per mental task, for each participant. A red cell indicates the exclusion of this participant in the analysis.

Participants	Mental tasks							Total (/336)
	ROT	WORD	SING	SUB	NAV	FACE	MOTOR	
S01	48	47	48	47	47	48	48	333
S02	30	21	25	21	30	27	30	184
S03	47	47	48	47	48	47	48	332
S04	45	48	46	47	47	46	47	326
S05	43	46	37	40	45	33	30	274
S06	43	45	44	40	43	40	42	297
S07	48	47	46	48	48	47	45	329
S08	48	35	30	48	43	26	24	254
S09	43	48	47	48	46	44	44	320
S10	48	43	31	47	42	39	43	293
S11	48	48	47	48	48	48	48	335
S12	14	16	13	16	1	16	16	92
Total	505	491	462	497	488	461	465	3369

subjects and across task pairs, the evolution of kappa across time, and the feature selection results are described.

4.2.1 EEG only

The peak classification performance of classifiers trained on one-second window EEG features, for each subject and pair of tasks, is shown in Table 4.3. The peak classification performance is defined as the

Table 4.3: EEG-only peak classification κ for each subject and task pair, when using features extracted from one-second windows. Task pairs and participants are shown in descending order of average κ, from top to bottom and left to right, respectively. κ values between 0.4 and 0.7 (satisfactory performance) are highlighted in light blue, whereas kappa values greater than 0.7 (high performance) are highlighted in dark blue. The average peak time with standard deviation is also included for each task pair.

	S10	S03	S06	S05	S07	S01	S11	S09	S04	Average	Peak time (s)
ROT-MI	0.93	0.90	0.95	0.80	0.87	0.73	0.79	0.77	0.63	0.82	3.5 ± 2.1
ROT-FACE	0.84	0.91	0.95	0.92	0.85	0.81	0.71	0.77	0.56	0.82	4.3 ± 2.7
ROT-SING	1.00	0.85	0.82	0.88	0.87	0.81	0.73	0.54	0.54	0.78	2.6 ± 1.1
SUB-MI	0.91	0.91	0.90	0.80	0.81	0.64	0.73	0.85	0.47	0.78	3.9 ± 4.1
SING-SUB	1.00	0.75	0.83	0.87	0.72	0.79	0.73	0.68	0.48	0.76	3.3 ± 2.3
SUB-FACE	0.69	0.85	0.93	0.89	0.71	0.81	0.71	0.67	0.40	0.74	3.7 ± 3.0
ROT-NAV	0.98	0.77	0.72	0.64	0.85	0.73	0.58	0.57	0.66	0.72	4.3 ± 1.0
ROT-WORD	0.96	0.74	0.61	0.89	0.71	0.58	0.69	0.58	0.53	0.70	3.5 ± 3.7
WORD-MI	0.95	0.56	0.89	0.57	0.74	0.60	0.69	0.65	0.43	0.68	5.1 ± 4.3
WORD-FACE	0.73	0.72	0.91	0.76	0.72	0.62	0.56	0.45	0.42	0.66	5.6 ± 4.2
NAV-MI	0.77	0.88	0.81	0.67	0.68	0.60	0.62	0.49	0.36	0.65	1.6 ± 2.3
SING-NAV	0.81	0.65	0.68	0.78	0.68	0.73	0.64	0.31	0.40	0.63	3.9 ± 5.2
SUB-NAV	0.86	0.60	0.50	0.77	0.69	0.62	0.50	0.68	0.45	0.63	6.7 ± 4.2
WORD-SING	0.92	0.66	0.82	0.51	0.70	0.71	0.60	0.34	0.38	0.63	4.2 ± 5.0
WORD-SUB	0.91	0.68	0.62	0.77	0.47	0.40	0.44	0.73	0.45	0.61	4.3 ± 4.4
NAV-FACE	0.43	0.79	0.83	0.82	0.56	0.41	0.48	0.31	0.31	0.55	1.8 ± 2.2
ROT-SUB	0.77	0.62	0.38	0.34	0.65	0.68	0.52	0.49	0.46	0.55	4.8 ± 2.8
WORD-NAV	0.74	0.56	0.41	0.43	0.62	0.47	0.50	0.36	0.37	0.50	8.5 ± 6.8
SING-MI	0.78	0.56	0.51	0.40	0.39	0.29	0.37	0.45	0.46	0.47	4.2 ± 5.4
FACE-MI	0.66	0.56	0.49	0.34	0.35	0.25	0.44	0.36	0.53	0.44	5.1 ± 6.7
SING-FACE	0.66	0.52	0.45	0.37	0.35	0.33	0.30	0.34	0.28	0.40	5.9 ± 5.5
Average	0.82	0.72	0.71	0.68	0.67	0.60	0.59	0.54	0.46	0.64	4.3 ± 4.2

highest $kappa$ obtained across the 15 seconds of task execution for windows of a specific length. All subjects achieved satisfactory or high performance for most task pairs, and with an average κ greater than 0.4. High performance was achieved for at least four different task pairs in all subjects but S04, and three subjects (S10, S03 and S06) showed an average peak κ greater than 0.7. A perfect classification score of $\kappa = 1$ was achieved for S10 in the ROT-SING and SING-SUB pairs.

The seven best performing task pairs were combinations of a brain-teaser and an imagery task: either ROT or SUB with SING, FACE, MI or NAV. For instance, the ROT-MI pair led to the highest average κ of 0.82. However, pairs of dynamic and static imagery tasks, including SING, FACE, NAV and MI, showed a consistently lower κ. For instance, the SING-FACE pair led to the lowest average κ of 0.4. The average EEG peak time across task pairs is of 4.3 seconds after stimulus onset with a high standard deviation of 4.2.

Table 4.4: EEG-only peak classification κ for each task pair, when using features extracted from one-second windows. Tasks are listed in descending order of average κ. The diagonal represents the average κ obtained for all pairs involving each specific task.

	ROT	SUB	MI	WORD	NAV	SING	FACE
ROT	0.73	0.55	0.82	0.70	0.72	0.78	0.82
SUB	0.55	0.68	0.78	0.61	0.63	0.76	0.74
MI	0.82	0.78	0.64	0.68	0.65	0.47	0.44
WORD	0.70	0.61	0.68	0.63	0.50	0.63	0.66
NAV	0.72	0.63	0.65	0.50	0.61	0.63	0.55
SING	0.78	0.76	0.47	0.63	0.63	0.61	0.40
FACE	0.82	0.74	0.44	0.66	0.55	0.40	0.60

Table 4.5: Top 10 EEG features for the five best task pairs, using the average absolute values of the SVM weights as ranking metric. This table uses the models trained on the one-second window occuring three to four seconds after stimulus onset, which corresponds to the average peak time for EEG classification (see Table 4.3).

	ROT-MI	ROT-FACE	ROT-SING	SUB-MI	SING-SUB
1	pwr_high_alpha_O2	pwr_high_alpha_O2	pwr_t4_30_P8	pwr_high_alpha_PO7	pwr_alpha/beta_PO8
2	pwr_low_alpha_CP6	pwr_t4_30_P10	pwr_t4_30_O2	pwr_alpha/beta_Iz	pwr_total_PO7
3	pwr_high_alpha_CP6	pwr_alpha/beta_O1	pwr_high_alpha_PO8	pwr_total_PO8	pwr_high_alpha_Iz
4	pwr_high_alpha_O1	pwr_high_alpha_PO7	pwr_t4_30_P10	pwr_theta/beta_O2	pwr_total_PO8
5	pwr_t4_30_P8	pwr_low_beta_P7	pwr_alpha/beta_PO8	pwr_theta/beta_P9	pwr_total_P7
6	pwr_alpha/beta_P6	pwr_theta/beta_O2	pwr_theta/beta_O1	pwr_alpha/beta_P6	pwr_t4_30_P9
7	pwr_low_beta_PO7	pwr_low_beta_PO7	pwr_low_beta_P7	pwr_total_P9	pwr_low_beta_TP7
8	pwr_theta/beta_PO8	pwr_high_alpha_O1	pwr_high_alpha_Iz	pwr_t4_30_F6	pwr_low_alpha_PO3
9	pwr_alpha/beta_Iz	pwr_theta/beta_Iz	pwr_t4_30_Pz	pwr_high_beta_CP1	pwr_high_alpha_O2
10	pwr_high_alpha_CP5	pwr_high_alpha_Iz	pwr_alpha/beta_Iz	pwr_high_alpha_POz	pwr_low_beta_PO7

We further summarize the results of Table 4.3 on a task-wise basis in Table 4.4. Viewed this way, it is clear that ROT and SUB tasks are the most useful overall, especially when paired with a passive imagery task. MI is the third best task, due to its performance when paired with brain-teasers.

The average classification κ over subjects, computed across one-second windows, is shown in Figure 4.1 for the six best task pairs. κ values increase quickly after stimulus onset, and reach a peak around four seconds later. The performance then decreases gradually over the remaining 11 seconds of the trial, and returns to chance level three seconds after the end of the task.

The l_1-SVM algorithm selected a minimum of zero[2] and a maximum of 12 EEG features, with a median of seven features. The evolution of the number of selected features is shown in Figure 4.2 for each participant to validate the behaviour of the l_1 feature selection procedure. During the pre- and post-epoch baseline, only around one or two features are selected. This number increases quickly after stimulus onset, and remains stable across the 15 seconds of the epoch. It is interesting to note that S04, who obtained the lowest average peak κ, also consistently had the lowest number of selected features.

[2]The case where zero features are selected simply means that all the SVM weights are zero, and so that the classifier always outputs the same decision, dependent on its bias b.

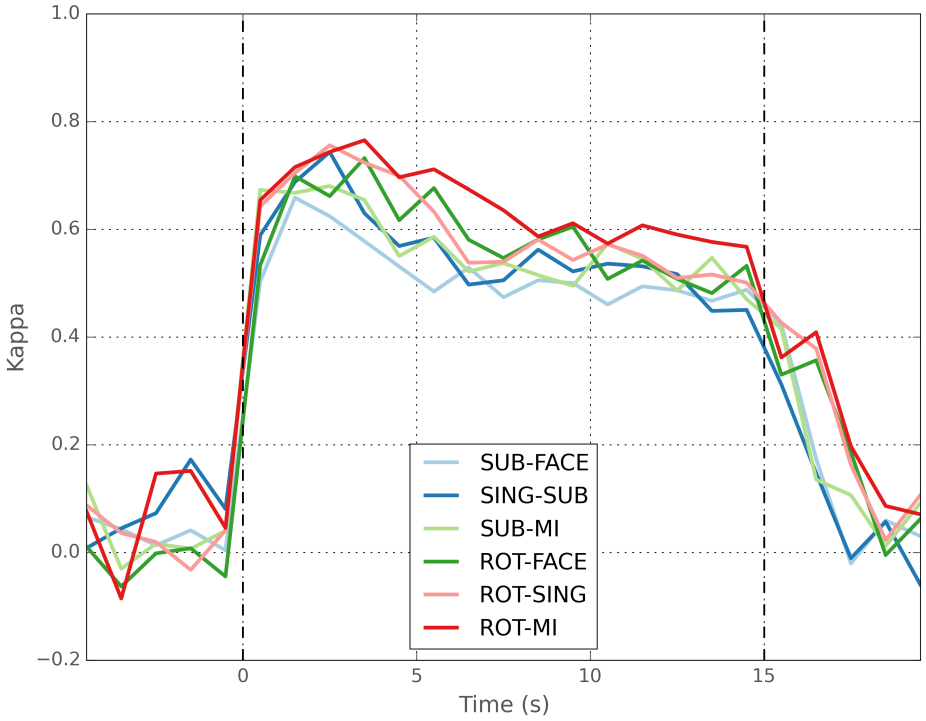

Figure 4.1: EEG-only classification κ over non-overlapping one-second windows for the six best task pairs. The classification κ obtained with one-second windows was averaged over participants for each task pair.

The 10 most important features following the approach described in Section 3.7 are listed in Table 4.5 for the five task pairs that produced the highest performance. α- and β-related features showed strong importance, while θ and wide-band features were typically not highly ranked. Most of these features were extracted from electrodes located in the parietal and occipital regions. To support this observation, the average absolute values of the SVM weights are visualized on a topographical map of the head to show the importance of each channel according to the trained classifier (see Figure 4.3). In almost every case, channels in the occipital and parietal regions showed the highest feature importance. High performance pairs (first three rows in Figure 4.3) all displayed maximum feature importance for channels at the back of the head. A pattern of high importance for channels PO7 and PO8 can be noticed in many pairs (i.e., ROT-FACE, SUB-FACE, ROT-NAV, SING-SUB and WORD-NAV).

4.2.2 NIRS only

The peak classification performance of classifiers trained on one-second window NIRS features, for each subject and pair of tasks, is shown in Table 4.6. The best overall κ is again obtained by S10, with 13 different pairs of tasks reaching high performance. S07, S04 and S11 respectively ranked fifth, last and seventh in the EEG-only analysis, all achieved an average κ above 0.50. On the other hand, S03, who

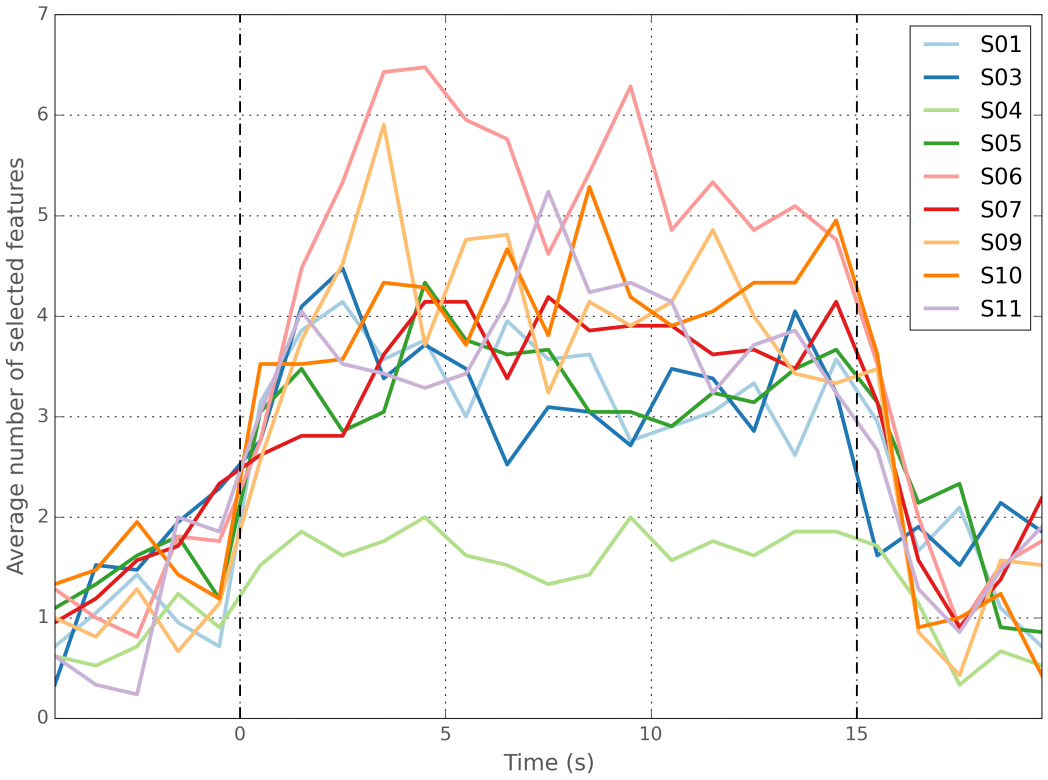

Figure 4.2: Average number of selected EEG features over non-overlapping one-second windows for all participants.

ranked second in the EEG-only analysis, was ranked second to last in the NIRS-only case. The average peak time across task pairs is 11.5 seconds, and is again very variable.

Fifteen pairs of tasks were classified with average κ above 0.4, but none achieved high performance on average. The best performing pairs were again a combination of a brain-teaser and an imagery task (ROT-MI, ROT-SING, ROT-FACE, SUB-MI, ROT-WORD). Pairs combining two imagery tasks, such as SING-MI, NAV-FACE, FACE-MI and SING-FACE, were constantly classified with the lowest average κ. As in the EEG-only case, ROT-MI achieved the highest performance, while FACE-MI and SING-FACE achieved the lowest. Again, we further summarize the results of Table 4.6 on a task-wise basis in Table 4.7. ROT and SUB tasks were the most useful overall when paired with passive imagery tasks, whereas pairs of imagery tasks are all under the 0.4 κ threshold.

The average classification κ over subjects, computed across one-second windows, is shown in Figure 4.4 for the six best task pairs. The κ values of the best task pairs oscillated slightly above $\kappa = 0$ during the baseline and the first five seconds of the trial. κ values then started rising up approximately five seconds after stimulus onset, and reached a plateau around six seconds later, which is consistent with the expected hemodynamic response delay [9]. Performance decreased after approximately eight seconds, and did not return to pre-epoch baseline levels in the five seconds after the end of the trial.

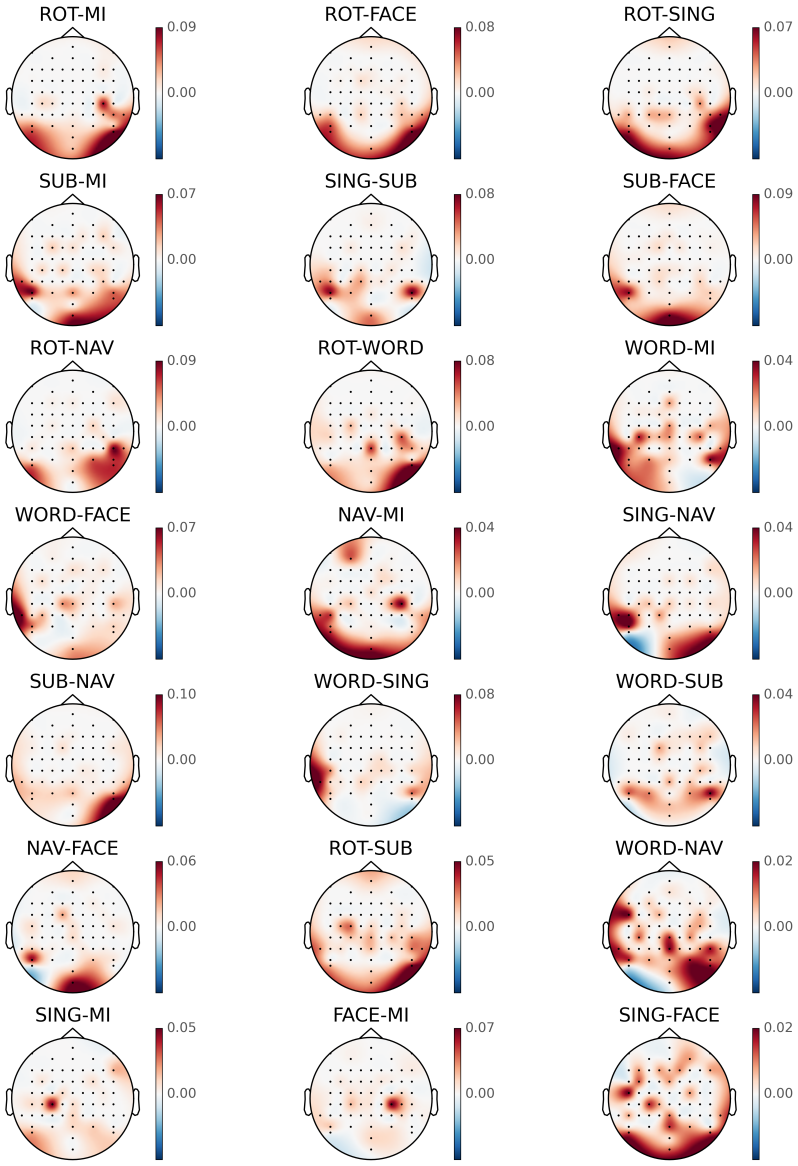

Figure 4.3: Topographical maps of the average absolute value of the SVM weights, when trained on EEG features only. Darker regions are those that are more important when classifying each task pair. This figure uses the models trained on the one-second window occurring three to four seconds after stimulus onset, which corresponds to the average peak time for EEG classification (see Table 4.3). The pairs are plotted in descending order of average κ (left to right, and top to bottom) as presented in Table 4.3.

When applied to NIRS features, the l_1-SVM algorithm selected a minimum of zero and a maximum of 11 features, with a median of three features. The evolution of the number of selected features is shown for each participant in Figure 4.5. Only around one feature was selected during the pre-epoch baseline and up to five seconds after stimulus onset, after which this number increased and remained more stable until the end of the task and the post-epoch baseline. S10, who obtained the highest average peak κ also consistently had the highest number of selected NIRS features.

Table 4.6: NIRS-only peak classification κ values for each subject and task pair, when using features extracted from one-second windows. Task pairs and participants are shown in descending order of average κ, from top to bottom and left to right, respectively. κ values between 0.4 and 0.7 (satisfactory performance) are highlighted in light blue, whereas κ values greater than 0.7 (high performance) are shown in dark blue. The average peak time with standard deviation is also included for each task pair.

	S10	S07	S04	S11	S09	S06	S01	S03	S05	Average	Peak time (s)
ROT-MI	0.82	0.72	0.71	0.60	0.65	0.65	0.58	0.56	0.30	0.62	11.6 ± 3.2
ROT-SING	0.77	0.68	0.80	0.66	0.49	0.50	0.60	0.52	0.40	0.60	11.8 ± 3.0
ROT-FACE	0.82	0.52	0.85	0.62	0.49	0.66	0.73	0.26	0.46	0.60	12.8 ± 2.4
SUB-MI	0.89	0.61	0.55	0.75	0.74	0.58	0.56	0.43	0.24	0.59	10.7 ± 6.1
ROT-WORD	0.85	0.68	0.74	0.60	0.66	0.52	0.58	0.32	0.30	0.58	13.2 ± 4.1
SING-SUB	0.70	0.66	0.61	0.62	0.71	0.43	0.54	0.60	0.39	0.58	13.8 ± 2.8
SUB-FACE	0.81	0.62	0.74	0.67	0.67	0.47	0.58	0.32	0.37	0.58	8.7 ± 7.4
ROT-NAV	0.76	0.71	0.80	0.50	0.46	0.70	0.52	0.39	0.36	0.58	11.8 ± 2.6
SUB-NAV	0.86	0.44	0.47	0.67	0.66	0.40	0.55	0.41	0.39	0.54	12.5 ± 3.0
WORD-SING	0.76	0.83	0.51	0.64	0.30	0.37	0.52	0.45	0.34	0.52	13.8 ± 4.6
WORD-SUB	0.65	0.56	0.73	0.69	0.56	0.28	0.51	0.26	0.42	0.52	13.7 ± 2.6
WORD-FACE	0.80	0.66	0.58	0.77	0.07	0.44	0.26	0.36	0.60	0.50	14.6 ± 2.9
WORD-NAV	0.84	0.75	0.43	0.56	0.36	0.32	0.28	0.29	0.30	0.46	13.5 ± 4.5
WORD-MI	0.74	0.69	0.53	0.69	0.25	0.38	0.26	0.16	0.38	0.45	12.9 ± 3.9
ROT-SUB	0.43	0.44	0.40	0.52	0.32	0.57	0.29	0.38	0.57	0.44	12.7 ± 4.6
SING-NAV	0.31	0.57	0.42	0.28	0.37	0.29	0.41	0.56	0.34	0.40	8.3 ± 7.5
NAV-MI	0.63	0.23	0.30	0.33	0.33	0.29	0.26	0.33	0.24	0.33	8.8 ± 7.6
NAV-FACE	0.55	0.24	0.42	0.31	0.31	0.26	0.22	0.20	0.27	0.31	10.2 ± 8.5
SING-MI	0.37	0.41	0.33	0.35	0.22	0.33	0.29	0.42	0.05	0.31	7.7 ± 6.2
SING-FACE	0.42	0.46	0.20	0.07	0.23	0.36	0.06	0.41	0.20	0.27	9.6 ± 6.0
FACE-MI	0.39	0.33	0.42	0.23	0.23	0.20	0.29	0.09	0.16	0.26	8.8 ± 4.3
Average	0.68	0.56	0.55	0.53	0.43	0.43	0.42	0.37	0.34	0.48	11.5 ± 5.2

The 10 most important features are listed in Table 4.8 for the five best task pairs. Sources 13 and 14 (right parietal cortex), and sources 15 and 16 (right and left prefrontal cortex) led to the highest ranking scores. HbR features generally showed more importance, followed by HbO features, whereas HbT features were included only four times in the 10 most important features for the best five task pairs. To further identify regions of interest, the average absolute values of the SVM weights are visualized on a topographical map of the head to show the importance of each channel according to the trained classifiers (see Figure 4.6). Channels in the prefrontal region showed consistently high importance for pairs including ROT and SUB tasks, whereas channels in the left prefrontal region showed consistent high feature importance in pairs that include the WORD task.

Table 4.7: NIRS-only peak classification κ for each task pair. Tasks are listed in descending order of average κ. The diagonal represents the average κ obtained for all task pairs using one specific task.

	ROT	SUB	WORD	SING	NAV	MI	FACE
ROT	0.57	0.44	0.58	0.60	0.58	0.62	0.60
SUB	0.44	0.54	0.52	0.58	0.54	0.59	0.58
WORD	0.58	0.52	0.51	0.52	0.46	0.45	0.50
SING	0.60	0.58	0.52	0.45	0.40	0.31	0.27
NAV	0.58	0.54	0.46	0.40	0.43	0.33	0.31
MI	0.62	0.59	0.45	0.31	0.33	0.43	0.26
FACE	0.60	0.58	0.50	0.27	0.31	0.26	0.42

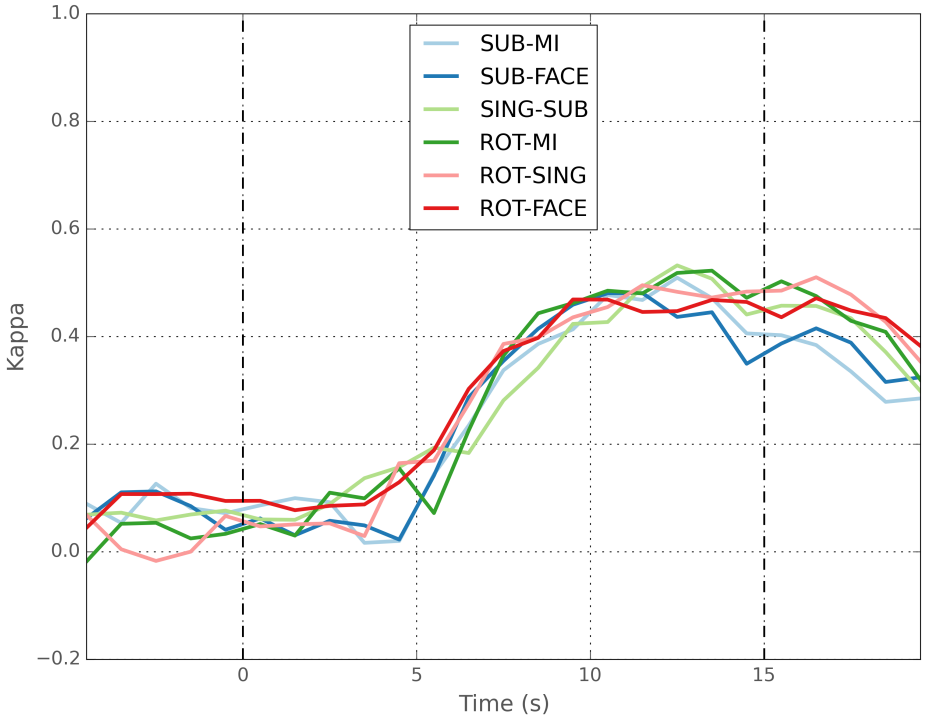

Figure 4.4: NIRS-only classification κ over non-overlapping one-second windows for the six best task pairs. The classification κ obtained with one-second windows was averaged over participants for each task pair.

4.2.3 Fusion of signals

The peak classification κ values obtained for models trained on concatenated EEG and NIRS features are shown in Table 4.9. The three best pairs of tasks remained the same as for EEG-only classification (ROT-FACE, ROT-MI and ROT-SING), as did the three worst pairs (SING-MI, FACE-MI, SING-FACE). Again, most of the highest ranked task pairs were a combination of a brain-teaser and an imagery task. Moreover, the task-wise κ values shown in Table 4.10 exhibited the same ranking as for EEG-

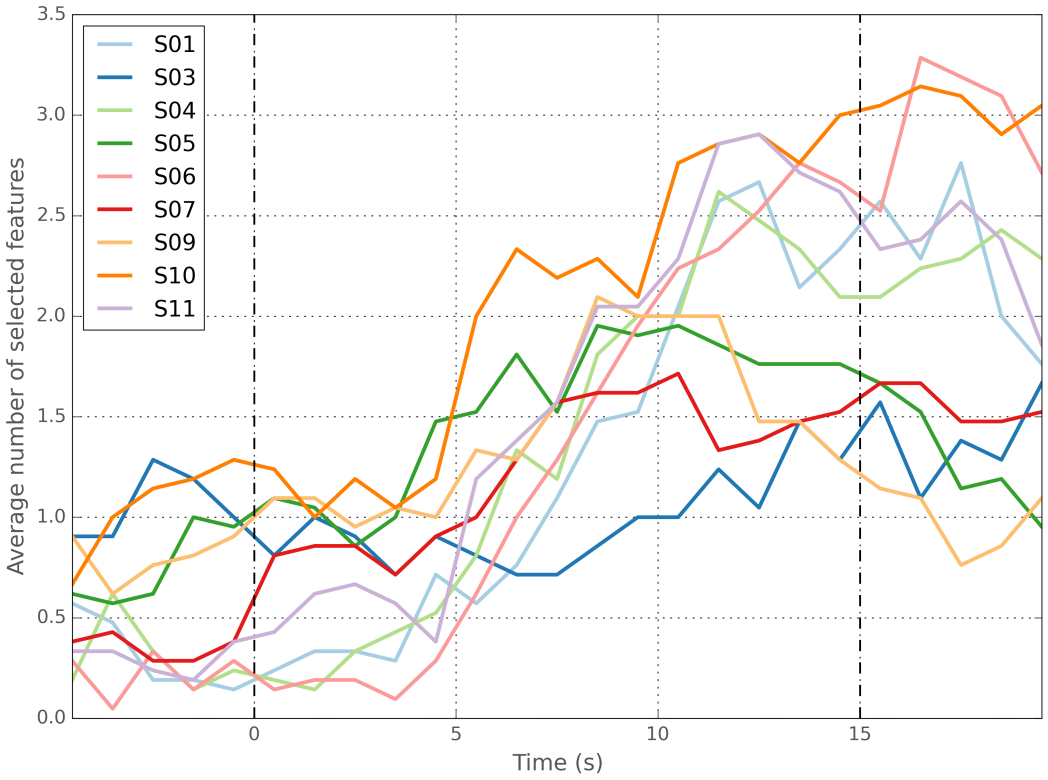

Figure 4.5: Average number of selected NIRS features over non-overlapping one-second windows for all participants.

Table 4.8: Top 10 NIRS features for the five best task pairs, using the average absolute values of the SVM weights as ranking metric. This table uses the models trained on the one-second window happening 11 to 12 seconds after stimulus onset, which corresponds to the average peak time for NIRS classification (see Table 4.6).

	ROT-MI	ROT-SING	ROT-FACE	SUB-MI	ROT-WORD
1	mean_S15-D1_HbO	mean_S15-D1_HbO	mean_S15-D1_HbO	mean_S14-D22_HbR	mean_S15-D5_HbO
2	mean_S10-D17_HbR	mean_S1-D2_HbT	mean_S10-D14_HbO	mean_S15-D1_HbR	mean_S13-D24_HbR
3	mean_S1-D2_HbT	mean_S6-D9_HbR	mean_S3-D6_HbR	mean_S15-D5_HbT	mean_S16-D17_HbR
4	mean_S13-D24_HbR	mean_S13-D22_HbR	mean_S10-D17_HbO	mean_S15-D1_HbO	mean_S13-D22_HbO
5	mean_S14-D24_HbR	mean_S14-D24_HbR	mean_S10-D17_HbR	mean_S1-D1_HbO	mean_S13-D24_HbO
6	mean_S13-D22_HbR	mean_S6-D6_HbR	mean_S15-D1_HbT	mean_S2-D5_HbR	mean_S15-D1_HbO
7	mean_S1-D2_HbO	mean_S4-D7_HbR	mean_S1-D13_HbT	mean_S10-D14_HbR	mean_S15-D1_HbT
8	mean_frM_HbO	mean_S10-D17_HbO	mean_S6-D6_HbR	mean_S10-D14_HbO	mean_S16-D13_HbR
9	mean_S6-D6_HbR	mean_S11-D19_HbR	mean_S4-D7_HbR	mean_S16-D13_HbO	mean_S1-D2_HbO
10	mean_S3-D6_HbR	mean_S15-D2_HbO	mean_S1-D2_HbT	mean_S9-D1_HbR	mean_S5-D7_HbO

only classification, with the exception of MI which achieved better performance in general than WORD. Subject-wise, the five best participants (S10, S06, S03, S07 and S05) were the same as for EEG-only classification. The other five subjects saw their ranking change by one or two positions. On the other hand, the peak time increased to an average of 6.4 seconds.

The average classification κ over subjects, computed across one-second windows, is shown in Figure 4.7 for the six best task pairs. Two peaks can be observed: first, at three seconds after trial onset, corresponding to the peak κ obtained with EEG features alone; and second, at around 11 seconds after

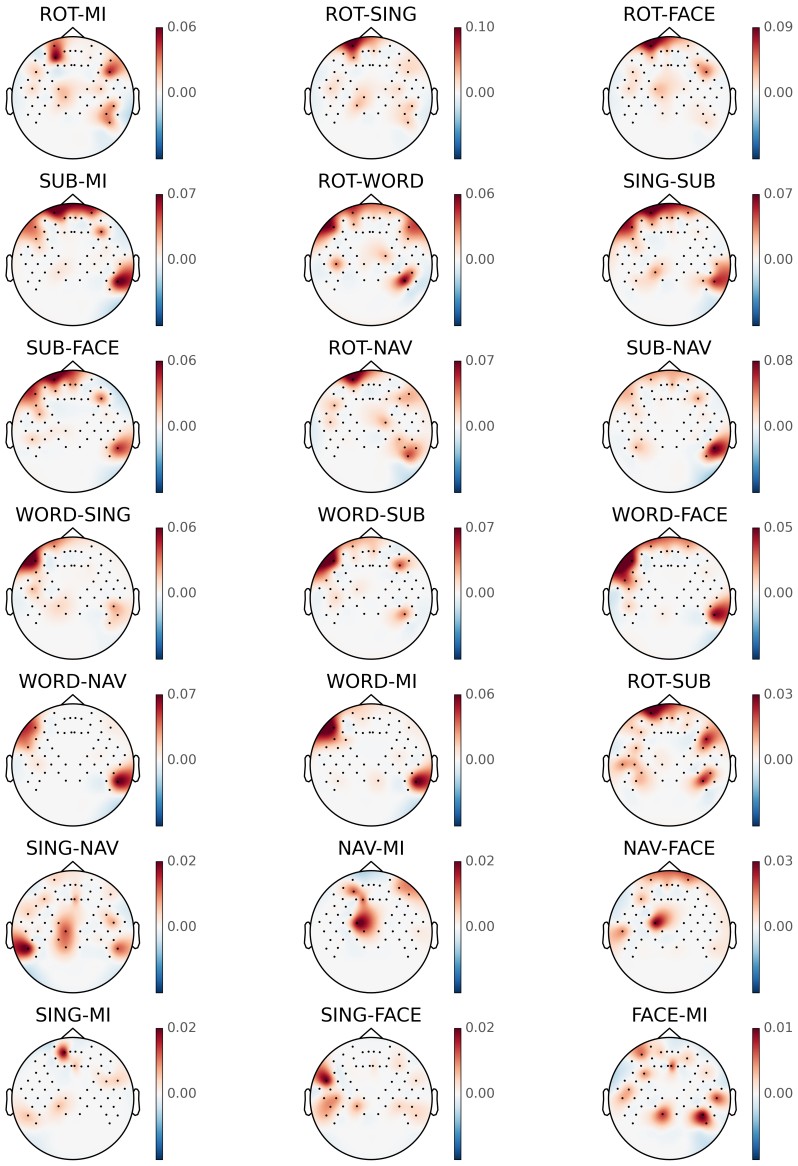

Figure 4.6: Topographical maps of the average absolute value of the SVM weights, when trained on NIRS features only. Darker regions are those that are more important when classifying each task pair. This figure uses the models trained on the one-second window occuring 11 to 12 seconds after stimulus onset, which corresponds to the average peak time for NIRS classification (see Table 4.6). The pairs are plotted in descending order of average κ (from left to right and top to bottom) as presented in Table 4.6.

trial onset, corresponding to the peak κ obtained with NIRS features alone. More specifically, when comparing these values to the ones obtained with EEG alone, we see that the improvement seems small in the first eight seconds of the task (as reported in Table 4.9), but is more noticeable in the last five seconds of the task, where more than one NIRS feature was originally selected (see Figure 4.5). Since the goal of this analysis is to explore the ways in which the fusion of EEG and NIRS can be useful, the rest of this section will be focused on the last five seconds of the 15-s tasks.

Table 4.9: EEG-NIRS fusion peak classification κ values for each subject and task pair, when using features extracted from one-second windows. Task pairs and participants are shown in descending order of average κ, from top to bottom and left to right, respectively. κ values between 0.4 and 0.7 (satisfactory performance) are highlighted in light blue, whereas κ values greater than 0.7 (high performance) are shown in dark blue. The average peak time with standard deviation are also included for each task pair.

	S10	S06	S03	S07	S05	S11	S01	S04	S09	Average	Peak time (s)
ROT-FACE	0.84	0.95	0.91	0.85	0.92	0.73	0.81	0.85	0.77	0.85	5.3 ± 3.8
ROT-MI	0.93	0.95	0.90	0.87	0.80	0.79	0.73	0.72	0.77	0.83	4.6 ± 4.0
ROT-SING	1.00	0.82	0.85	0.91	0.88	0.73	0.81	0.83	0.57	0.82	5.5 ± 4.9
SUB-MI	0.91	0.90	0.91	0.81	0.80	0.81	0.64	0.55	0.85	0.80	6.3 ± 5.6
SING-SUB	1.00	0.83	0.81	0.79	0.87	0.73	0.79	0.61	0.74	0.80	6.2 ± 5.0
SUB-FACE	0.81	0.93	0.85	0.71	0.89	0.75	0.81	0.74	0.67	0.80	6.9 ± 4.6
ROT-NAV	0.98	0.72	0.77	0.85	0.64	0.58	0.73	0.83	0.57	0.74	6.6 ± 4.8
ROT-WORD	0.96	0.62	0.74	0.71	0.89	0.62	0.60	0.74	0.69	0.73	7.7 ± 6.2
WORD-MI	0.95	0.89	0.56	0.80	0.56	0.83	0.60	0.53	0.65	0.71	7.3 ± 5.6
WORD-FACE	0.78	0.91	0.72	0.72	0.76	0.77	0.62	0.53	0.45	0.70	7.8 ± 4.2
WORD-SUB	0.91	0.62	0.68	0.58	0.77	0.69	0.47	0.71	0.73	0.68	8.5 ± 6.7
WORD-SING	0.92	0.82	0.66	0.78	0.53	0.62	0.71	0.58	0.34	0.66	7.8 ± 5.8
NAV-MI	0.84	0.81	0.88	0.68	0.69	0.65	0.60	0.40	0.42	0.66	2.2 ± 2.8
SUB-NAV	0.91	0.47	0.60	0.69	0.77	0.73	0.62	0.45	0.68	0.66	7.9 ± 4.4
SING-NAV	0.81	0.68	0.65	0.68	0.78	0.64	0.73	0.40	0.31	0.63	1.7 ± 3.6
ROT-SUB	0.77	0.52	0.62	0.65	0.57	0.54	0.68	0.46	0.52	0.59	7.5 ± 5.2
NAV-FACE	0.56	0.83	0.79	0.56	0.82	0.48	0.43	0.42	0.24	0.57	4.1 ± 5.1
WORD-NAV	0.83	0.48	0.56	0.75	0.43	0.52	0.45	0.41	0.44	0.54	9.7 ± 5.9
SING-MI	0.78	0.53	0.56	0.44	0.40	0.37	0.29	0.46	0.45	0.48	6.2 ± 4.1
FACE-MI	0.66	0.44	0.56	0.30	0.33	0.44	0.21	0.53	0.36	0.43	7.6 ± 7.1
SING-FACE	0.66	0.43	0.52	0.46	0.43	0.26	0.33	0.28	0.34	0.41	6.2 ± 5.8
Average	0.85	0.72	0.72	0.69	0.69	0.63	0.60	0.57	0.55	0.67	6.4 ± 5.2

To get a better idea of how the feature fusion of EEG and NIRS impacted the classification performance, Table 4.11 shows the increase in peak κ obtained by adding NIRS to EEG features when considering the one-second window spanning from 11 to 12 seconds after stimulus onset, which was found to be the optimal NIRS window (see Table 4.6). The average increase was of 0.17, while the highest increase was of 0.96, achieved for S07 on task pair SUB-MI. Oppositely, the largest decrease was of -0.29, for S09 in task pair WORD-NAV. Average performance gains were the highest for S04 ($\Delta\kappa = 0.27$), S07 ($\Delta\kappa = 0.23$) and S06 ($\Delta\kappa = 0.19$). While S04 and S07 achieved good NIRS-only performance (third and second best, respectively), S06 obtained below average performance. S04 was also the most poorly ranked in EEG-only. S10, who achieved the best performance in both unimodal conditions, showed an average increase in the multimodal context, but still suffered a decrease for a few task pairs such as SING-MI and NAV-MI. S05 and S01, who obtained the third and second to last lowest κ increases, were amongst the most poorly ranked in NIRS-only classification (ninth, eighth and sixth, respectively). However, S09, who benefited the least from fusion, had only obtained a slightly below average NIRS-only performance.

Table 4.10: EEG-NIRS fusion peak classification κ for each task pair. Tasks are listed in descending order of average κ. The diagonal represents the average κ obtained for all task pairs using one specific task.

	ROT	SUB	WORD	MI	NAV	SING	FACE
ROT	0.76	0.59	0.73	0.83	0.74	0.82	0.85
SUB	0.59	0.72	0.68	0.80	0.66	0.80	0.80
WORD	0.73	0.68	0.67	0.71	0.54	0.66	0.70
MI	0.83	0.80	0.71	0.65	0.66	0.48	0.43
NAV	0.74	0.66	0.54	0.66	0.63	0.63	0.57
SING	0.82	0.80	0.66	0.48	0.63	0.63	0.41
FACE	0.85	0.80	0.70	0.43	0.57	0.41	0.62

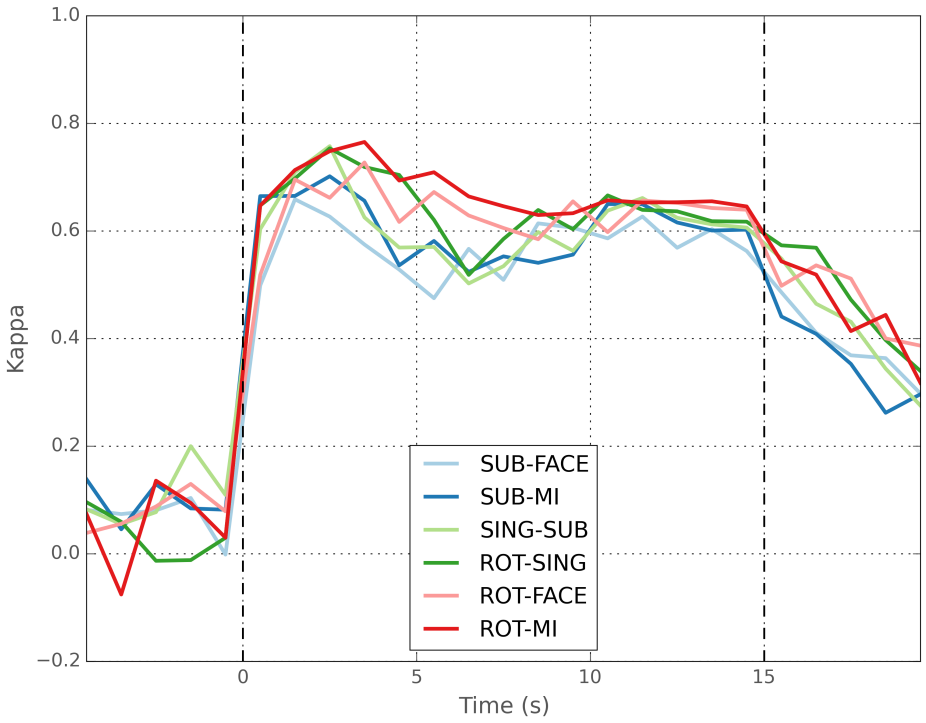

Figure 4.7: EEG and NIRS classification κ over non-overlapping one-second windows for the six best task pairs. The classification κ obtained with one-second windows was averaged over participants for each task pair.

The three task pairs that achieved the best average increases included the subtraction task: SUB-MI ($\Delta\kappa = 0.41$), SING-SUB ($\Delta\kappa = 0.34$) and SUB-FACE ($\Delta\kappa = 0.30$), three tasks that also achieved good NIRS-only performance. Only one pair (FACE-MI) saw an average decrease in peak κ with the addition of NIRS, while two pairs (NAV-MI and SING-FACE) saw a negligible increase of 0.01.

The l_1-SVM algorithm selected a minimum of zero and a maximum of 15 features with a median of seven features when trained on the concatenated EEG-NIRS dataset. The evolution of the number of selected features is shown in Figure 4.8 for each participant. The behaviour is almost identical to that

Table 4.11: Change in peak κ when adding NIRS features to EEG features, for the models trained on the one-second window spanning 11 to 12 seconds after trial onset. Task pairs and participants are shown in descending order of average $\Delta\kappa$, from top to bottom and left to right, respectively. $\Delta\kappa$ values between 0.4 and 0.7 are highlighted in light blue, values greater than 0.7 are shown in dark blue and values below 0 are shown in light red.

	S04	S07	S06	S11	S10	S03	S05	S01	S09	Average
SUB-MI	0.55	0.96	0.88	0.44	-0.11	0.17	0.50	0.07	0.21	0.41
SING-SUB	0.56	0.90	0.02	0.08	0.87	0.17	0.43	-0.04	0.11	0.34
SUB-FACE	0.66	0.53	0.52	0.15	-0.03	0.38	0.19	0.08	0.26	0.30
WORD-MI	0.63	0.30	0.92	0.35	0.12	-0.15	0.25	0.04	-0.00	0.27
WORD-SUB	0.50	0.31	0.22	0.58	-0.00	-0.09	0.26	0.36	0.04	0.24
ROT-WORD	0.24	0.33	-0.02	0.33	0.05	0.45	0.20	0.32	0.16	0.23
WORD-SING	0.02	0.18	0.09	0.32	0.65	0.19	-0.01	0.50	0.11	0.23
WORD-FACE	0.38	0.64	0.79	0.12	0.09	0.23	-0.19	0.02	-0.05	0.23
ROT-SUB	0.13	0.13	0.13	0.13	0.77	0.06	0.48	0.11	0.00	0.21
SUB-NAV	0.43	0.10	0.33	0.06	0.46	0.48	0.05	-0.04	-0.00	0.21
NAV-FACE	0.40	0.09	0.25	0.40	0.25	0.11	-0.16	-0.04	0.22	0.17
SING-NAV	0.15	0.45	-0.14	0.06	-0.01	-0.06	0.49	0.31	0.02	0.14
ROT-FACE	0.66	-0.10	0.10	-0.21	0.07	0.02	0.10	0.21	0.24	0.12
WORD-NAV	0.04	0.29	0.26	0.31	0.26	0.11	0.08	-0.02	-0.29	0.12
ROT-SING	0.29	-0.08	0.00	0.23	0.05	0.04	-0.03	0.29	0.18	0.11
ROT-MI	-0.02	0.07	0.05	0.38	0.00	0.11	-0.03	0.06	-0.02	0.07
ROT-NAV	0.35	0.12	-0.05	-0.06	0.03	0.08	0.07	0.21	-0.16	0.07
SING-MI	0.00	-0.27	-0.12	0.05	-0.04	0.52	0.31	-0.19	0.05	0.04
NAV-MI	-0.17	0.18	-0.07	0.33	-0.46	0.52	-0.11	-0.18	0.04	0.01
SING-FACE	0.04	-0.19	-0.13	0.06	0.59	0.05	-0.25	-0.15	0.05	0.01
FACE-MI	-0.10	-0.09	-0.02	-0.17	0.14	0.17	-0.04	0.02	-0.05	-0.02
Average	0.27	0.23	0.19	0.19	0.18	0.17	0.12	0.09	0.05	0.17

exhibited by EEG-only models, until around 11 seconds after stimulus onset, where a slight increase in the number of selected features can be seen. This effect is particularly apparent for S04, who benefited the most from the fusion.

The 10 most important features per task are listed in Table 4.13 for the six best task pairs. Most of the previously chosen features are selected again (Tables 4.5 and 4.8). Figure 4.9 shows the relative importance of each subtype of feature (band power and chromophore type) for the same six task pairs. We see that α-related features were generally the most important, followed by β-related features. On the other hand, the θ, 4-30 Hz and 0.1-100 Hz bands were usually less important. In NIRS, HbR-related features showed the strongest importance (while still inferior to α-related features), followed closely by HbO-related features. HbT did not show much importance. In some participants (S03, S07 and S11), HbR-related features were consistently more important than HbO-related ones, while it was the opposite for S04 and S09 (results not shown).

Table 4.12: Change in peak κ for each task pair when adding NIRS to EEG features. Tasks are listed in descending order of average κ. The diagonal represents the average κ obtained for all task pairs using one specific task. $\Delta\kappa$ values between 0.4 and 0.7 are highlighted in light blue and values below 0 are shown in light red.

	SUB	WORD	SING	FACE	ROT	MI	NAV
SUB	0.29	0.24	0.34	0.30	0.21	0.41	0.21
WORD	0.24	0.22	0.23	0.23	0.23	0.27	0.12
SING	0.34	0.23	0.14	0.01	0.11	0.04	0.14
FACE	0.30	0.23	0.01	0.14	0.12	-0.02	0.17
ROT	0.21	0.23	0.11	0.12	0.13	0.07	0.07
MI	0.41	0.27	0.04	-0.02	0.07	0.13	0.01
NAV	0.21	0.12	0.14	0.17	0.07	0.01	0.12

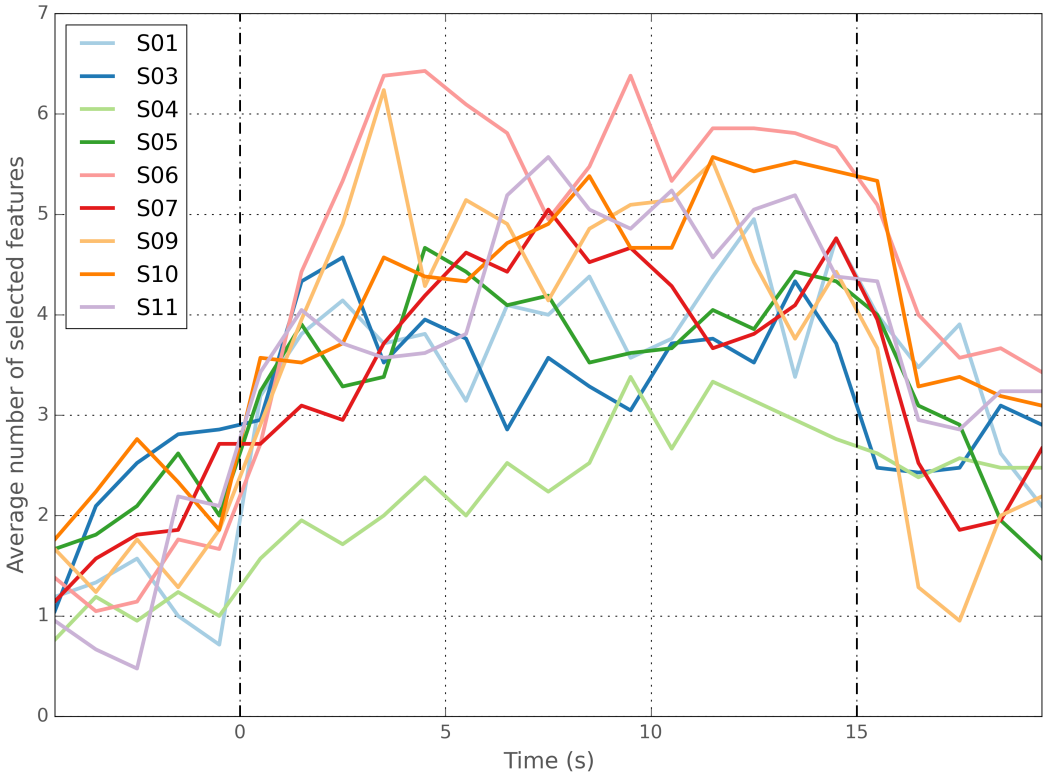

Figure 4.8: Average number of selected fusion features over non-overlapping one-second windows for all participants.

4.2.4 Effect of time window selection on classification performance

The effect of using different window sizes for extracting features on classification performance is shown in Figure 4.10. Longer time windows generally produced higher peak κ for EEG and fusion, with the exception of 15-s windows that scored lower than 10-s windows. This was not the case for NIRS however, which saw its average peak κ steadily decrease down to 0.5 when 15-s windows were used. Feature fusion yielded higher increases for shorter than for longer windows.

Table 4.13: Top 10 fusion features for the five best task pairs, using the average absolute values of the SVM weights as ranking metric. This table uses the models trained on the one-second window happening 11 to 12 seconds after stimulus onset, which corresponds to the average peak time for NIRS classification (see Table 4.6).

	ROT-FACE	ROT-MI	ROT-SING	SUB-MI	SING-SUB
1	mean_S15-D1_HbO	pwr_theta/beta_O2	mean_S15-D1_HbO	mean_S14-D22_HbR	pwr_high_alpha_O2
2	pwr_t4_30_O2	pwr_high_alpha_O1	pwr_high_alpha_O2	pwr_high_alpha_PO7	pwr_t4_30_P9
3	pwr_theta/beta_O2	pwr_t4_30_O2	pwr_t4_30_P9	pwr_t4_30_O2	pwr_high_alpha_P10
4	pwr_theta/beta_PO7	pwr_t4_30_PO8	pwr_t4_30_O2	pwr_high_alpha_O2	mean_S15-D1_HbO
5	pwr_high_alpha_PO8	mean_S14-D24_HbR	pwr_theta/beta_O1	pwr_high_alpha_PO8	pwr_low_alpha_P9
6	mean_S10-D14_HbO	mean_S1-D2_HbO	pwr_alpha/beta_PO8	mean_S1-D1_HbO	mean_S14-D22_HbR
7	pwr_t4_30_P9	pwr_t4_30_P9	pwr_low_alpha_O2	mean_S2-D5_HbR	pwr_t4_30_CP2
8	mean_S10-D17_HbO	pwr_alpha/beta_PO8	pwr_low_beta_PO7	pwr_alpha/beta_Iz	mean_S15-D5_HbT
9	pwr_high_alpha_O1	mean_frM_HbO	pwr_t4_30_P10	mean_S15-D5_HbT	pwr_high_alpha_P7
10	pwr_alpha/beta_Iz	mean_S13-D22_HbR	mean_S14-D24_HbR	pwr_theta/beta_O2	mean_S1-D13_HbO

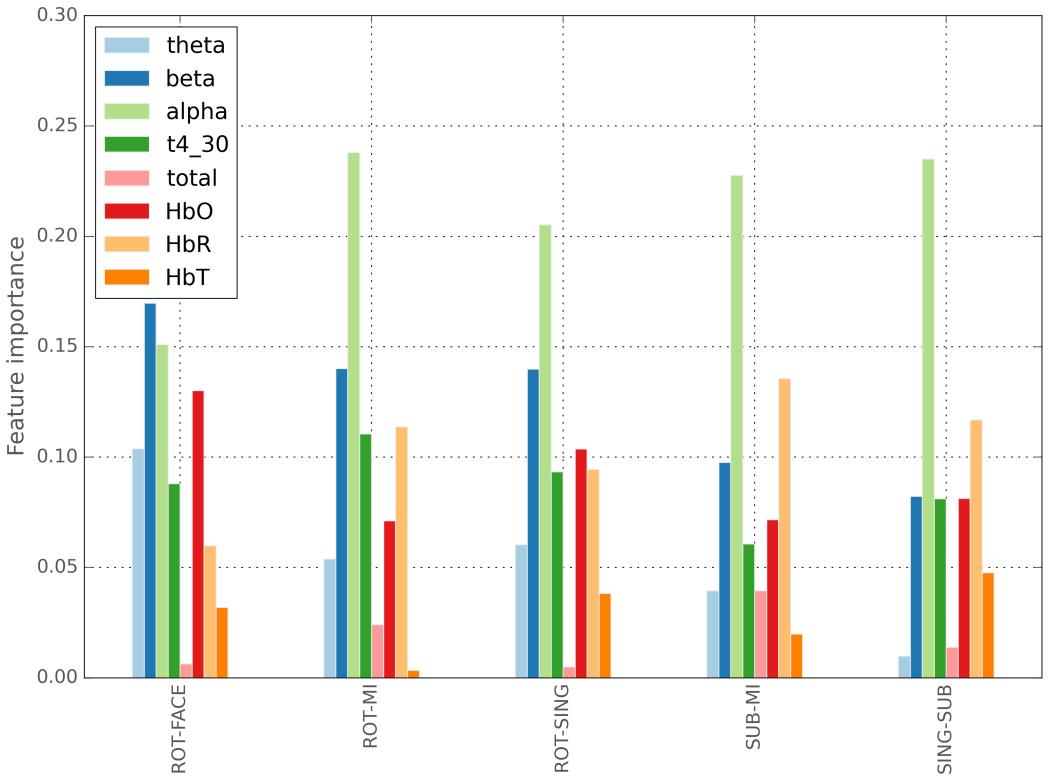

Figure 4.9: Importance of each EEG and NIRS feature subtype: EEG band powers (θ, α, β, 4-30 Hz and total spectrum (0.1-100 Hz)) and chromophore types (HbO, HbR and HbT). The histogram is based on the models trained on the one-second window occuring 11 to 12 seconds after stimulus onset, which corresponds to the average peak time for NIRS classification (see Table 4.6).

4.3 Neurophysiological data descriptive analysis

The neurophysiological data was analyzed to provide more insights into the mental processes involved in the tasks. Specifically, we focus our analysis on the spatial patterns of activation with ERS/ERS maps in EEG and chromophore maps in NIRS.

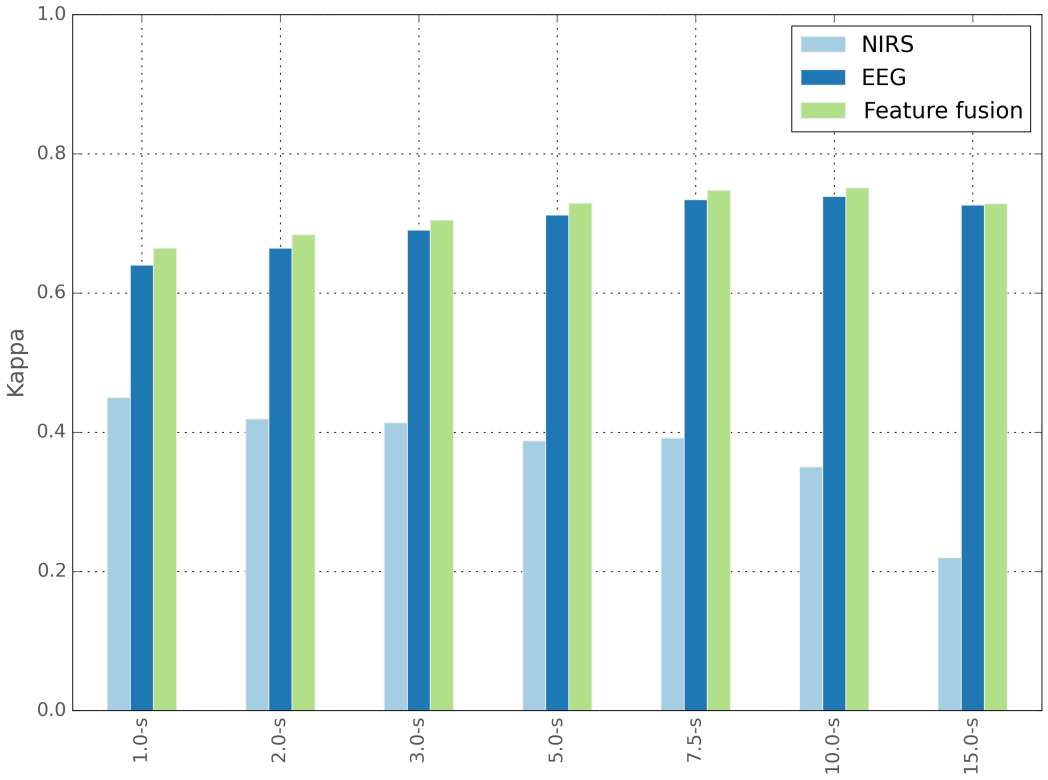

Figure 4.10: Average peak classification κ obtained for different window sizes with EEG, NIRS and EEG-NIRS fusion. The average peak κ across subjects and task pairs was computed for windows of 1, 2, 3, 5, 7.5, 10 and 15 seconds (see Table 3.2).

4.3.1 EEG

The ERD/ERS maps for each task in the low α, high α, low β and high β bands are shown in Figure 4.11. In mental rotation, high amplitude ERD patterns were measured in the occipital region for all four bands. A pattern of weak ERS was observed over the two motor cortices in all bands except low β. Word generation produced ERS patterns over the left temporal lobe in the low α and high α bands. Patterns elicited by mental subtraction showed consistently high ERD in the occipital lobe, especially around PO7 and PO8 (both amongst the most important features identified in Table 4.5). In mental singing, ERD patterns over the left hemisphere in the low and high β bands were observed, as well as predominant all-band ERS in the occipital region. Mental navigation led to left hemispheric ERD in the low α and low β bands, as well as left and right prefrontal ERD in the high β band. In motor imagery, left hemispheric ERD patterns were dominant in the first three bands, except for a bilateral high α ERD pattern in the prefrontal lobe, similar to the one observed for NAV but of lower intensity. Finally, face imagery led to ERD patterns in the frontal (again, similar to the ones observed for NAV and MI in the high β band) and temporal lobes, as well as high α power in the occipital lobes.

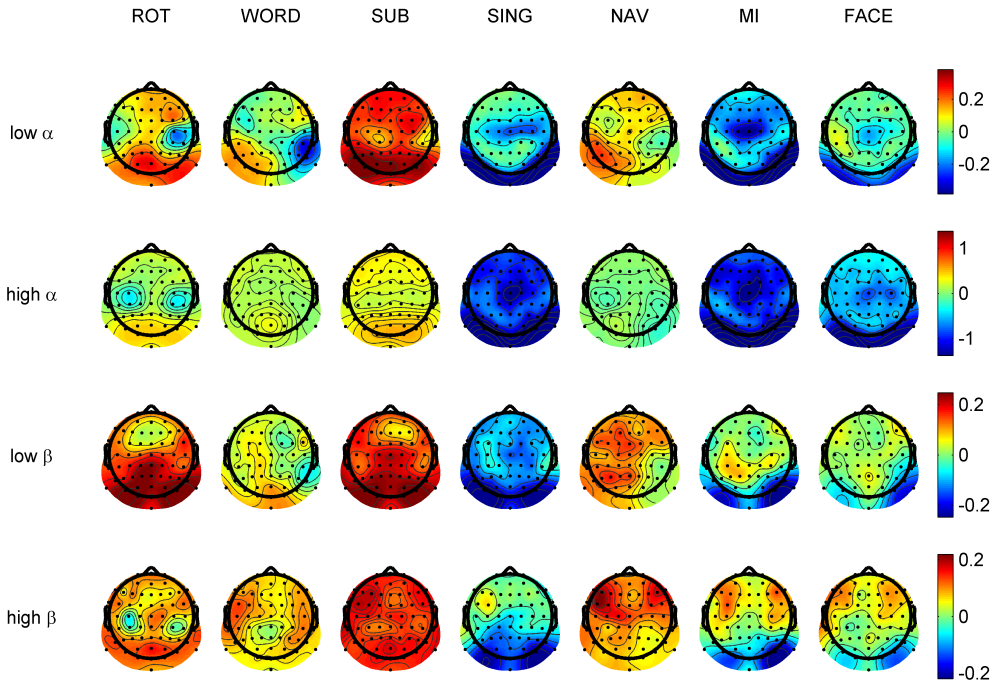

Figure 4.11: ERD/ERS maps for each task in the low α, high α, low β and high β bands. ERD/ERS values are computed using the inter-trial variance method [73] over the time window spanning 0.5 to two seconds after stimulus onset, using a baseline of -2 to 0 seconds before trial onset. Blue represents ERS while red represents ERD.

4.3.2 NIRS

Figure 4.12 shows the average HbO, HbR and HbT topographical maps for each task during the time window spanning 10 to 15 seconds after stimulus onset. Mental rotation led to a decrease in HbO and HbT over the frontal lobe, as well as an increase in HbR over the prefrontal lobe. In turn, word generation yielded an increase in HbO over the left temporal lobe and a decrease in HbR over the left and right temporal lobes. This was accompanied by a widespread increase in HbT over the posterior left hemisphere. In mental subtraction, an increase in HbO and HbT over both the right and left temporal lobes, as well as a decrease in the prefrontal region, was observed. Oppositely, HbR increased over the midline, but decreased in both temporal lobes. Mental singing provoked a subtle decrease in HbR over the temporal lobes, similar to WORD and SUB. The mental navigation task led to decreased HbR levels over both temporal lobes and increased HbO levels in the same regions. A large increase in HbO and HbT was observed over the left hemisphere for the motor imagery task, without any patterns of similar amplitude in HbR. Finally, face imagery led to an increase in the levels of all chromophores on the midline. A consistently high HbR pattern over the right centro-parietal region can also be seen in most tasks.

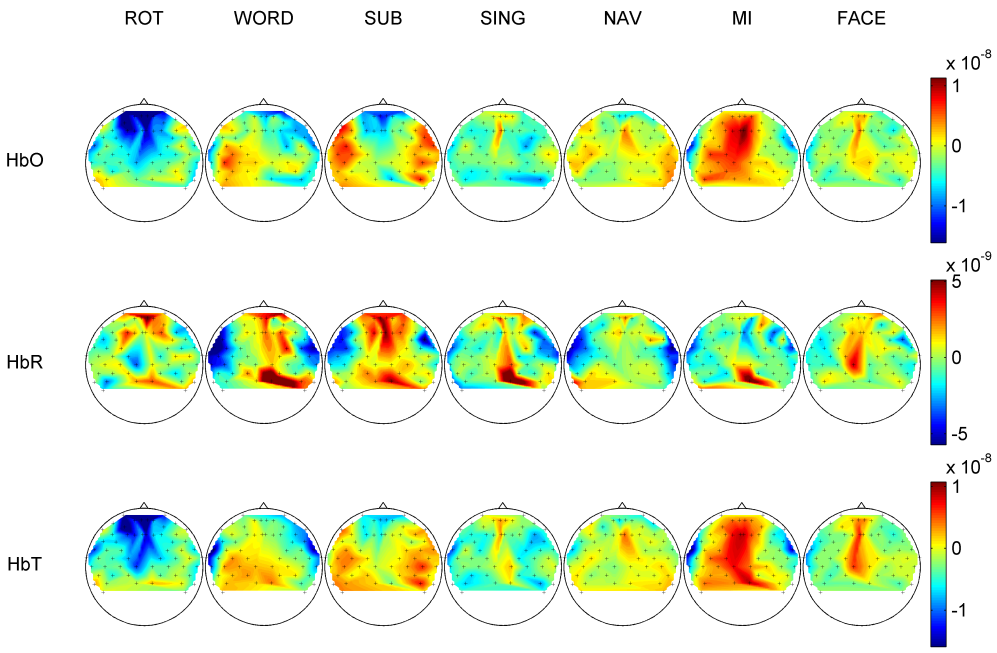

Figure 4.12: Average HbO, HbR and HbT maps for each task. Reported HbO, HbR and HbT values are normalized to their baseline values, and averaged across the time window spanning 10 to 15 seconds after stimulus onset. Red represents an increase in concentration of the chromophore, while blue represents a decrease.

4.4 Questionnaire results

The subjective ratings, collected at each recording session, are analyzed to assess the potential usability of the different mental tasks in a hBCI context. In a first step, the NASA TLX ratings and the ranking of tasks given by the participants are described. Second, a correlation analysis between the features and the ratings is performed to uncover features that could be used as markers to measure the subjective experience of participants.

4.4.1 NASA TLX ratings

The average ratings, as well as the task rankings, are shown in Figure 4.13. Overall, the SUB task induced the most mental and temporal load, effort, and frustration, led to the poorest perceived performance, and was the least preferred. Other brain-teasers (WORD and ROT) induced a medium load, effort and frustration, but were ranked high against other tasks in terms of preference. On the other hand, dynamic imagery tasks (SING, NAV and MI) induced a relative low load, did not require much effort, and were the least frustrating, while leading to high perceived performance, but mixed preference rankings. Finally,

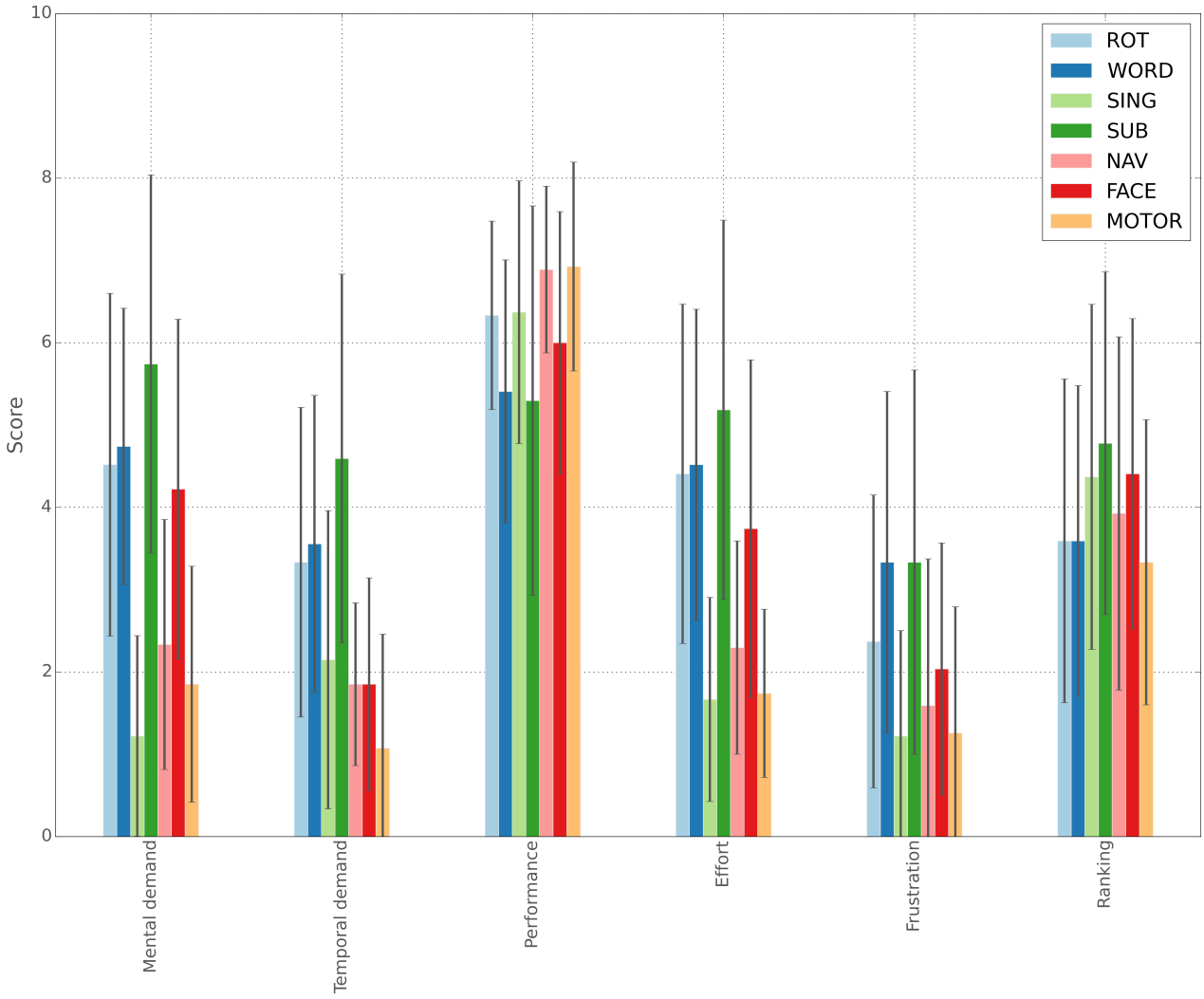

Figure 4.13: NASA TLX ratings and mental task ranking. Ratings and rankings are averaged over each subject and session. Error bars correspond to the standard deviation of each group of rating and mental task.

the only passive imagery task (FACE), required high mental load and effort but low temporal load, and led to slightly lower performance and higher frustration levels, while being ranked amongst the least preferred tasks.

4.4.2 Correlation analysis between neurophysiological features and subjective ratings

Features that are significantly correlated with the subjective ratings are shown in Tables 4.14, 4.15, 4.16, 4.17 and 4.18. The nomenclature used to name the extracted features is described in Appendix B. These results are now described in more detail for each subjective dimension.

Table 4.14: NIRS features significantly correlated with the mental demand rating.

Feature	ρ_{spear}
vall_cfL_HbR	-0.409914

Table 4.15: EEG and NIRS features significantly correlated with the temporal demand rating.

Feature	ρ_{spear}
alpha_m-delta_P7	-0.405477
alpha_m-delta_P9	-0.416606
delta_m-delta_PO7	0.427031
alpha_m-delta_PO7	-0.425669
diff_S1-D4_HbO	-0.488970
diff_S1-D13_HbO	-0.432533
diff_S16-D13_HbO	-0.416185
diff_S1-D4_HbT	-0.479567
diff_pfR_HbO	-0.411533
diff_frM_HbO	-0.419349
peak_S1-D4_HbO	-0.467170
peak_S1-D4_HbT	-0.490290
fit1a_S8-D18_HbO	-0.494393
fit1a_S9-D3_HbO	0.493041
fit1a_S10-D14_HbO	-0.417295
fit3a_S3-D7_HbO	-0.485993
fit3a_S4-D8_HbO	0.498769
fit3c_cfL_HbR	-0.407912

Mental demand

As shown in Table 4.14, only one feature is selected: the latency at which the HbR signal reached its lowest value in the left centro-frontal artificial channel. This feature was found to be negatively correlated with mental demand.

Temporal demand

For temporal demand, a total of 18 features were found to be significantly correlated with the subjective ratings (see Table 4.15). First, four amplitude modulation features over the left parietal and parieto-occipital regions in the α and δ bands were selected. Many NIRS features also showed significant correlations. Second, the amplitude difference between the first and second halves of the window led to seven significant features in the prefrontal lobe (midline and right) for HbO mainly, as well as peak latency in HbO and HbT in the same regions. Third, coefficients of the linear and cubic fits (in HbO mainly) were significant in many locations (central, prefrontal right and temporal left).

Table 4.16: EEG features significantly correlated with the performance rating.

Feature	ρ_{spear}
pwr_high_gamma/beta_Cz	-0.405137
alpha_m-theta_T7	0.403803
alpha_m-theta_CP3	0.459752
beta_m-delta_CP3	0.418309
beta_m-theta_CP3	0.430731
beta_m-delta_CP1	0.426476
beta_m-theta_CP1	0.434770
beta_m-beta_CP1	0.408494
beta_m-delta_CPz	0.406969
alpha_m-theta_C4	0.422127
alpha_m-theta_C6	0.437945
alpha_m-theta_CP4	0.405130
alpha_m-theta_CP2	0.454861
beta_m-delta_CP2	0.409816
beta_m-theta_CP2	0.409516

Table 4.17: NIRS features significantly correlated with the effort rating.

Feature	ρ_{spear}
vall_cfL_HbR	-0.401652

Performance

Only EEG features were selected in our correlation analysis for perceived performance: all except one were amplitude modulation features with positive correlation. Amongst them, the θ and δ modulations of α and β bands over the central and centro-parietal regions were selected the most often.

Effort

Only one feature was selected for the effort rating, which was the same as for mental demand (vall_cfL-HbR).

Frustration

Finally, one EEG and one NIRS features were selected for the frustration rating: the β-modulated β band over the C6 electrode, as well as the second order coefficient of the quadratic fit of HbR over the centro-parietal cortex.

Table 4.18: EEG and NIRS features significantly correlated with the frustration rating.

Feature	ρ_{spear}
beta_m-beta_C6	-0.415954
fit2a_S5-D9_HbR	0.413171

Chapter 5

Discussion

The results presented in Chapter 4 are discussed in the following sections. First, we discuss the optimal mental tasks and combinations of mental tasks that are the best candidates for an online BCI implementation, using EEG-only, NIRS-only and EEG-NIRS fusion. Second, we compare overall modality performances to identify the strengths and weaknesses of each proposed modality configuration. Third, the correlation analysis of questionnaire ratings and neurophysiological features is discussed. Finally, we highlight the main limitations of this study and provide suggestions to improve the results in future work.

5.1 Optimal mental task and combinations

As a fairly young field of research, Brain-Computer Interfacing still has a lot of room for improvement. For example, most research has focused on a few well explored paradigms such as SMR, ERP and SSEP (see Section 2.1.2), while alternatives such as non-motor mental tasks have been less extensively studied. However, to improve the usability and flexibility of BCIs, new paradigms have to be identified and validated. In this work, we studied seven different mental tasks (mental rotation, word generation, mental subtraction, mental singing, mental navigation, motor imagery and face imagery) from an electrophysiological and neurohemodynamic perspective in order to uncover the most promising contenders. In the following sections, we discuss the results obtained with each mental task individually and highlight its potential usability in an online BCI.

5.1.1 Mental rotation

Task pairs with mental rotation were always ranked amongst the best pairs, in all modality configurations (EEG-only, NIRS-only and EEG-NIRS fusion). Pairs including ROT and any task other than SUB yielded peak κ higher than average in each modality configuration. These results confirm the findings of previous

studies [36, 61], in which ROT was also amongst the best tasks. However, in our study ROT was performed in a slightly different way: participants were shown two figures and had to rotate one to evaluate if it was the same as the other one or a mirrored version, instead of simply imagining a single figure rotating. The introduction of a clear goal puts this task in the brain-teaser category with WORD and SUB, which were previously shown to yield high performance. However, the average increase in performance brought in by feature fusion was amongst the lowest of any tasks.

The features selected for the high performing pairs using ROT were consistently more important at the back of the head in EEG and in the left prefrontal region for NIRS. The ERD/ERS analysis in Figure 4.11 supports this observation: high levels of ERD were observed in the occipital regions in all four plotted bands, suggesting that these regions were recruited during mental rotation. Similar ERD patterns were found in [36] for the low β band between 0.5 to 2 seconds after the beginning of the task, with additional ERD in the prefrontal region. A distinct pattern over the two motor cortices can be seen in both α and β, suggesting motor imagery might have been used by some participants to help mentally rotate the L-shaped figures. The first findings are confirmed by a review looking at 32 fMRI and Positron Emission Tomography (PET) neuroimaging studies that concluded that the posterior occipital cortex was consistently activated during mental rotation [67]. In this review, the author also notes the activation of focused prefrontal cortex regions, and more precisely the left inferior frontal cortex for studies that encouraged motor simulation (that is, when participants were asked to imagine manipulating the objects to be rotated). Although this was not precisely the case here, participants were not given specific instructions as to how they should perform the rotation, and so some might have used this approach, explaining the strong HbO decrease in the prefrontal patterns found in NIRS.

Although ranked as third most demanding, frustrating and effort-inducing task, ROT was often amongst the three preferred tasks. Therefore, even though ROT might induce fatigue more rapidly than other tasks, its high performance in both NIRS and EEG and its good preference ranking make it an excellent candidate for a BCI.

5.1.2 Word generation

Word generation led to above average classification performance for all three modality configurations, and usually ranked third or fourth amongst tasks. These results are similar to those observed in EEG for WORD pairs in [36]. The fusion proved to be particularly useful for WORD, which achieved the second best increase in κ in a multimodal context. Indeed, all task pairs with WORD except WORD-NAV were in the top eight of largest absolute κ increase. This increase might in part be explained by the good signal quality often obtained on the contour of the head, where the cap was tighter and the optodes less obstructed by hair. Indeed, this consistently good signal quality across participants might have helped NIRS play a more important role in the fusion context.

The features selected in pairs using WORD were predominantly in the left temporal region in EEG, and in the left fronto-temporal region in NIRS. In terms of ERD/ERS, this effect was mostly noticeable in the low α and high β bands, in which the left temporal region undergoes desynchronization. It is also interesting to note that the low β pattern characterized by desynchronization in the occipital and central left regions is very similar to the one found in [36]. The temporal regions were also highlighted in NIRS, with strong HbO increase over the left temporal lobe and HbR decrease over both sides. Again, these findings are further confirmed by recent fMRI studies focused on anatomical regions supporting different aspects of language [83]. Indeed, in their review, Price *et al.* found that the left middle frontal cortex and the left pars opercularis (Broca's area) were consistently activated during word retrieval tasks. These brain areas overlap with the regions identified in our results.

Similarly to ROT, WORD ranked as the second most demanding, frustrating and effort-inducing task, but was still amongst the three preferred tasks. Contrary to ROT though, participants often rated their performance for the WORD task as low. This might be explained by the purely random selection of the first letter from which words had to be generated: some letters are rarely found at the beginning of a word (such as Z, X or Q), and so participants who were given these letters probably performed worse. The difficulty of the WORD task would need to be adjusted in a future implementation by limiting the selection of some rarer letters. Overall, word generation is a good candidate for a BCI thanks to its high performance in both NIRS and EEG and its good subjective evaluation.

5.1.3 Mental subtraction

Mental subtraction ranked second in terms of average classification performance for all three modality configurations. SUB is the task that benefited the most from the feature fusion, and the three largest increases in κ were all a combination of SUB and an imagery task (MI, SING and FACE). These results confirm the findings of [36] in EEG, but not of [61] in NIRS. Indeed, Hwang *et al.* (2014) found mental subtraction to be at par with other imagery tasks (SING, MI) in terms of how often it led to classification accuracies above 70%. However, the authors defined their SUB task as the successive subtraction of two "simple numbers" (suggesting one-digit numbers) from a three-digit number, which is simpler than the task used in our study. Moreover, a mental multiplication task (of two two-digit numbers), which was not used here, led to the best performance in their work. Since some fMRI studies have found the two operations to be similarly encoded in the brain (although some studies found differences) [192], and since the difficulty level of Hwang *et al.*'s multiplication task might be closer to our subtraction task, we hypothesize that they could have achieved similar performance with harder subtractions.

The features selected in pairs using the SUB task revealed consistent importance at electrode positions PO7 and PO8 in EEG, and in the prefrontal and right temporo-parietal regions in NIRS. This is also seen in the ERD/ERS patterns as a strong desynchronization in low and high α as well as low β in the occipito-parietal regions and part of the prefrontal regions. FMRI studies identified the involvement of the precuneus, located in the midline portion of the centro-parietal region, and of frontal regions as

commonly activated during the execution of different arithmetic operations [84]. Interestingly, in the same study, Fehr *et al.* found a statistically significant increase in activity in the bilateral inferior parietal regions when comparing a complex subtraction task to a simple one, which could explain these clear patterns of importance for PO7 and PO8 features. As for the prefrontal patterns found in NIRS, they might be related to an increased working memory load, which would result in the activation of the dorsolateral prefrontal cortex [85]. The patterns of HbR decrease in both temporal lobes being very similar to the ones observed in WORD, we hypothesize participants might have used a speech-based strategy to perform the subtractions.

As a brain-teaser, SUB ranked as the most demanding, frustrating and effort-inducing task, making it the worst perceived for performance and the least preferred task. Before using SUB in a BCI, it might be useful therefore to further assess its optimal difficulty level to avoid tiring or frustrating the user. Additionally, alternative arithmetic operations such as mental multiplication might yield better results and should therefore be studied [61], again at optimal difficulty levels. In either case, SUB can definitely benefit from a EEG-NIRS fusion approach and is thus recommendable if such a system is already in use.

5.1.4 Mental singing

Task pairs using mental singing led to below average performance for all three modality configurations, but to the third highest average κ increase in fusion. Analogous results were obtained for EEG [36] and for NIRS [61] in previous studies comparing many mental tasks. Similarly, in a prefrontal NIRS study, Power *et al.* (2011) compared the performance of rest-SING and rest-SUB pairs and found that mental singing led to poorer performance on average [56].

It is difficult to determine if consistent brain regions were recruited in the mental singing task based on the feature importance analysis, as in both EEG and NIRS the topographical maps show varying patterns for pairs using SING. Indeed, in EEG, parieto-occipital (with ROT), bilateral parietal (with SUB), left parietal and right occipital (with NAV), left temporal (with SING), left motor cortex (with MI) and occipital (with FACE) all showed strong importance. Similar disparate patterns were found in NIRS. This high variability in importance of brain regions is most likely due to the low discriminability of the physiological processes behind mental singing, especially in the studied modalities, that are essentially limited to cortical structures. Indeed, if little information from singing could be decoded in EEG and NIRS, a classifier would focus on features provided by the second task to make its decision, and thus would not choose consistent features across pairs containing SING.

Neuroimaging studies of covert singing identified activations in the fronto-parietal regions (bilateral motor cortices and Broca's area) [86, 87]. This could explain the ERD patterns in the low and high β band located in the left hemisphere and the similar HbR patterns over the temporal and central lobes. These studies also provide insights into why mental singing did not produce distinguishable patterns: first, Gunji *et al.* (2007) used MEG to compare the oscillatory processes of humming, speaking, overt

and covert singing, and found that covert singing consistently recruited the least cerebral area [87]. A very small area of activation will make it harder for modalities such as EEG and NIRS to pick up relevant activity when restricted to a small number of sensors. Second, in their study of opera singers, Kleber *et al.* (2007) found significant activation of emotion-related deep brain structures such as the insula, the amygdala and the hippocampus [86]. Since in our study we specifically asked the participants to focus on the emotions produced by singing a song they personally chose, it is possible that these regions were activated but that the depth limitations of EEG and NIRS prevented these patterns from being measured. The strong α and β ERS visible in Figure 4.11 further supports this hypothesis. Indeed, the high increase in these oscillatory processes indicates that most cortical structures were not activated during singing. However, the ERS patterns being especially strong in the occipital region, it is possible that they are the result of low mental load-induced α power (see Section 5.5.1).

SING induced the lowest mental demand, effort and frustration, and was reasonably well ranked by participants. Since pairing SING with brain-teasers led to high classification performance, it seems reasonable to use it in a BCI context. Nevertheless, because of the apparent low discriminability of mental singing in EEG and NIRS, it would be important to first assess how it compares to baseline data: if it is not different enough from the baseline, this task could eventually provoke false negatives in a BCI using a no-control state. It is to be noted that this analysis was not performed here due to baseline periods being used for normalization of the features, thus limiting the significance of such an analysis on our dataset.

5.1.5 Mental navigation

Task pairs using mental navigation led to below average performance for all three modality configurations, making them very similar in terms of performance to pairs including mental singing, and yielded the lowest average κ increase in fusion. Similar results were obtained in [36] for EEG-only classification, whereas no previous NIRS results were found in the literature.

ERD/ERS patterns for mental navigation showed mostly left hemispheric activations, particularly in the low α and low β bands. These results are consistent with the low β patterns found in [36] above the left hemisphere, and are further supported by the results of an fMRI study looking at 16 subjects engaged in mental navigation of familiar places [88]. In their work, Ino *et al.* (2002) reported statistically significant activations in the left premotor area, the left angular gyrus (parietal lobe) and deeper brain structures such as the bilateral retrosplenial areas, parts of the hippocampus and the cerebellum [88]. These patterns are not easily discerned in the feature importance topographical maps, where different brain regions seem informative for each pair in both EEG and NIRS, as was the case for SING. However, Ino *et al.* used a more complex navigation task in which participants were instructed to imagine walking through Kyoto while counting the number of turns they made. Since both our study and [36] obtained poor performance with this task, we hypothesize that increasing the difficulty level might have helped produce stronger and clearer activations.

NAV ranked as the third least demanding, frustrating and effort-inducing task, and led to high perceived performance. Ranked as fourth favourite task on average, NAV, like SING, should be considered for implementation in a BCI mostly because it led to good performance when paired with brain-teasers. Increasing the difficulty level of the task might however lead to superior performance. Finally, its low κ increase in the multimodal case makes it less useful to use with a NIRS-EEG approach.

5.1.6 Motor imagery

Motor imagery is undoubtedly the most often used mental task in the BCI literature [14], providing a wealth of previous articles to compare our results to. Whereas MI was the second worst task in NIRS, it achieved third and fourth best performance in EEG only and EEG-NIRS configurations, respectively. MI was also the second worst task in terms of performance improvement when feature fusion was used. Although using a different processing and classification pipeline, Fazli *et al.* (2012) worked toward a similar hybrid EEG-NIRS BCI as in our study, but only evaluated the use of a left-hand vs. right-hand MI paradigm [17]. The authors found that EEG-NIRS fusion could improve the classification accuracy by 5% on average in 90% of subjects, which is hard to compare to our study because of methodological differences. However, we also noted a marked improvement across participants ($\Delta\kappa = 0.13$), even though this improvement was amongst the lowest in our seven tasks.

Although patterns induced by MI are most likely obscured by those of brain-teasers in most pairs (as was the case for SING and NAV), feature importance analysis shows the importance of left and right motor cortex features in NAV-MI, SING-MI and FACE-MI in EEG. This phenomenon seems to occur only for the NAV-MI pair in NIRS, but this might reveal in fact a pattern proper to NAV, as the NAV-FACE map shows similar feature importance. The ERD/ERS patterns in the first three bands showed activation of the motor cortices, with the left hemisphere being predominant, as expected from the right hand movement to be imagined. Classical BCI experiments have shown this pattern multiple times [89, 90], and the particular pattern in low β is very similar to the one reported in [36]. This is further supported by strong HbO and HbT increases over the left hemisphere. However, desynchronization of both the left and right frontal cortices is visible in the high β band, a phenomenon not reported in the aforementioned literature. Similar bilateral frontal ERD patterns were found for almost every task in high β, suggesting that there might have been an unexpected brain activity-inducing event affecting this band. As is the case for SING, strong ERS in the low and high α bands, predominantly in the occipital lobe, suggests that the lower mental load induced by this task might lead to drowsiness-related artefacts (see Section 5.5.1).

MI induced either the lowest or second lowest demand, effort and frustration, yielded the highest perceived performance, and was ranked as the most pleasant task on average. Pairing MI with brain-teasers led to high classification performance in our case, and was shown in the literature to yield good EEG and NIRS performance when paired with another imagery task (such as another MI task performed

with the other hand) [17]. We thus conclude that MI is a useful task that deserves its predominant place in the BCI literature.

5.1.7 Face imagery

Face imagery was the worst task in all modality configurations. Fusion improved the classification performance of FACE, but not differently from most other tasks. These results are in line with those of Friedrich *et al.* for EEG [36], whereas no reference for NIRS-based classification of face imagery could be found.

The main brain structure known to be activated specifically in response to face stimuli is the fusiform gyrus, found in the posterior part of the cortex [91]. In addition, in reaction to famous faces, activation of subsets of the inferior occipital gyri, lateral fusiform gyri, superior temporal sulcus and amygdala was identified in [91]. The fact that the main structure recruited (fusiform gyrus) is farther from the cortical layers accessible with EEG and NIRS could certainly account for the difficulty in classifying FACE tasks. This could also explain the disparate feature importance topographical patterns observed in both EEG and NIRS.

The ERD/ERS pattern observed in the low β band was similar to the one reported in [36]: desynchronization of the central and frontal lobes, with synchronization of the temporal lobes. Moreover, α bands showed distinctly high power in the occipital lobes, which could again be explained by some low mental load-induced drowsiness. This is especially relevant since some of the brain structures mentioned above are located in the parietal and occipital brain regions.

Although inducing low temporal demand, the FACE task was almost at par with brain-teasers for mental demand and effort, suggesting participants found it difficult to imagine the face of their closest friend. Falling amongst the three least preferred tasks, FACE is however not a good candidate for a BCI, as it produced consistently low unimodal and multimodal performance, and was generally not particularly appealing to users.

5.1.8 Best task pairs

Based on these results, we conclude that out of the 21 task pairs studied in our study combinations of a brain-teaser (ROT, SUB or WORD) and an imagery task (MI, FACE, NAV, SING) are most likely to yield good performance. This confirms the previous results of [36] and [61]. In the case where a NIRS-EEG system is available for the implementation of the BCI, our results show that pairs based on SUB and WORD might perform even better than in unimodal conditions.

5.2 Evaluation of fusion performance in a realistic context

A few recent studies aimed at combining EEG and NIRS in a hybrid BCI implementation (see Section 2.3). Our study followed a similar procedure as [17], in which both EEG and NIRS were used to classify the same task pairs, following the hypothesis that the two modalities can provide complementary information on the underlying neurophysiological processes. However, Fazli *et al.* (2012) only assessed the use of motor imagery and execution, without exploring other mental tasks, and used a slightly different methodology. Whereas similar NIRS features were extracted, EEG features were computed using a CSP filter, which we did not apply for reasons stated in Section 3.5.3. The authors also used a separate LDA classifier for EEG, HbO and HbR, and a meta-LDA to fuse the decisions of each previous classifier. Moreover, their analysis was based on classification accuracy, whereas ours is based on Cohen's κ. These methodological differences make it difficult to compare our results. Nonetheless, both their study and ours showed it is possible to improve unimodal classification performance by combining EEG and NIRS features.

In our case, a few methodological points would have to be approached differently in a fully online implementation. First, the analysis of the effect of fusion on classification performance was done in two parts. We started by comparing the peak κ achieved by fusion to the peak κ of EEG alone, and saw small increases on the order of $\Delta\kappa = 0.02$. By looking at the evolution of performance across time (Figures 4.1, 4.4 and 4.7), we later found NIRS provided additional information around 11 seconds into the task. We thus chose to report the performance increase for the one-second window between 11 and 12 seconds after stimulus onset to highlight the added value of a fusion approach. This approach would have to be adapted when considering a real-time BCI. For example, many NIRS studies extracted the same feature over windows of many different sizes and combined them as different features of the same mental task instance [56–58, 61, 62]. While this approach can hamper classifier performance by increasing the dimensionality of the input, it also provides more information and thus can lead to better accuracy.

Another critical point highlighted in many of the aforementioned studies is the delay introduced by the HRF in NIRS. Indeed, in our study, we found it took around 11 seconds for NIRS features to yield peak performance. This is a long time to wait when trying to operate a BCI in real-time, making a unimodal NIRS-BCI poorly usable in most contexts. In turn, some participants saw their EEG performance decrease steadily across a trial, probably due to them being bored or tired, reducing the quality of their mental imagery and solving skills. However, by combining NIRS with EEG, we could reach satisfactory performance in the first seconds after stimulus onset, while still benefiting from increased performance later in the trial.

Despite these limitations, several applications could benefit from the peculiarities of our hBCI paradigm. For example, one could design a BCI where a decision has to be sustained over a few seconds before an output is given to ensure a certain level of certainty. Our system would be useful to

recognize a specific mental task over longer periods of time, while also providing increased classification performance.

Another example of how the fusion of EEG and NIRS features might be useful was shown in our analysis for participant S04. Indeed, this participant had the lowest EEG-only performance, and did not reach peak κ above 0.7 for a single task pair when using EEG features only. However, this participant's NIRS-only performance proved much better, and when EEG and NIRS were combined S04 benefited from the largest performance increase across participants and achieved high performance in seven task pairs. This pattern was observed for many participants: usually, one modality led to better performance than the other, and combining both EEG and NIRS allowed improvements in many task pairs. For example, this means a BCI user with poor EEG-only control would still be able to achieve high performance.

Finally, a major challenge in offline brain-computer interfacing studies like this one is the evaluation of the proposed approach in real-work contexts [57]. For instance, the setup cost and complexity, the required processing time and power, need to be taken into account, especially when two different modalities are used at once.

5.3 Effect of time window selection

The effect of time window selection on the classification performance, as shown in Figure 4.10, revealed expected patterns in EEG and fusion, but not in NIRS. Indeed, longer time windows produced higher classification performance in both EEG and fusion. This is expected as longer windows will typically incorporate more information about the temporal processes occurring during the task and thus will lead to better discriminable features. Specifically, in EEG, we expect the quality of the extracted features to increase with longer windows, since the frequency resolution of the spectrum increases with longer windows. Similarly, feature fusion was expected to follow a similar trend as EEG, but with some added performance provided by NIRS.

However, shorter NIRS windows surprisingly led to better performance than longer ones. This could be explained by the fact that average chromophore levels are less informative as the windows grow. Indeed, longer windows, especially the ones that span both the low performance period (between 0 to 11 seconds after stimulus onset) and the high performance period after that, will lose information because of the averaging of non-relevant and relevant information. For windows of less than 7.5 seconds, however, this is more likely explained by the loss of information brought in by local averaging. In either case, more investigations have to be done to explain this phenomenon.

Because both the BCI response time and the classification performance grow with longer windows, choosing a window length thus becomes a trade-off between the responsiveness and the accuracy of the system. In our case, we saw that one-second windows could lead to good classification performance

in most subjects, while benefiting from the highest increase in average κ produced by fusion. In BCI applications where delays are acceptable, it might nonetheless be more useful to adopt longer windows.

5.4 Correlation analysis between neurophysiological features and subjective ratings

The analysis of the correlation between the extracted neurophysiological features and the ratings collected using the NASA TLX questionnaire was conducted to unveil potential markers of subjective experience that could be used in a BCI context. This supplementary information would be useful in implementing passive BCIs that could adapt the difficulty of the mental tasks to the user's mental state, for example. We discuss the significance of these results in this section.

First of all, our experiment was not designed to provide an unbiased evaluation of subjective ratings. Indeed, although the ratings were collected three times for each participant and mental task, we did not vary the difficulty level of each task: this means it is likely, if the ratings do not vary noticeably within each mental task, that the results reflect inherent differences between tasks and not just the differences in mental load induced in participants. Therefore, in order to validate these findings, it would be necessary to conduct additional studies where the inter-task differences are controlled for. However, by cross-validating our findings with the task-wise results discussed in the previous section, it is possible to infer if these results were likely caused by the inter-task differences or some other physiological processes.

The mental demand and effort dimensions, being correlated with an ρ_{spear} of 0.79, both showed a negative correlation with HbR levels in the left centro-frontal region. The opposite effect was observed however in the chromophore level analysis (Figure 4.12), where high effort- and mental demand-inducing tasks (ROT, WORD, SUB and FACE) actually led to higher levels of HbR. As noted above, the recruitment of this region might be explained by the increase in working memory load caused by these tasks, yielding an activation of the dorsolateral prefrontal cortex [85].

Temporal demand was found to negatively correlate with amplitude modulation α features over the parieto-occipital region, and with many HbO features over the prefrontal region. The former pattern matches with the parieto-occipital ERS seen in the α and β bands for the low temporal demand tasks (SING, MI and FACE), but not for NAV, which also produced low temporal demand but did not show these ERS patterns. The correlations are significant despite the opposite effect seen in NAV due to the high ERS seen in the three other tasks. As for NIRS, the decreased HbO over prefrontal regions in ROT, SUB and WORD (and the increased HbO in the other tasks, especially MI) could explain these correlations. Therefore, it seems that both these significant correlations observed for temporal demand can be explained by the inter-task differences.

For perceived performance, many amplitude modulation α and β features showed positive correlation over centro-parietal regions. Low performance tasks (WORD, SUB and FACE), however, showed very different ERD/ERS patterns, as was the case for high performance tasks (ROT, SING, NAV and MI). The effect captured by the correlation analysis might thus be very specific to the amplitude-modulated α and β bands (as supported by the fact that none of the classical log-power features were selected). Moreover, an fMRI study on nine participants showed a fronto-parietal BOLD increase in trials when a correct answer was given in a visual working memory task [92], which overlaps with the selected brain region in our analysis. Therefore, the perceived effect might indeed be related to task performance, although more work has to be done to validate this hypothesis.

Finally, two features were selected for the frustration dimension over the central and centro-parietal regions. In [93], the effect of frustration was assessed in fMRI, and was found to recruit the amygdala, a part of the midbrain, deep parts of the cortex (insula) and the prefrontal cortex. The lack of correspondence between our findings and the literature indicates that more work has to be done to correctly assess the effect of the frustration dimension.

Overall, our exploratory correlation analysis of neurophysiological features and subjective ratings showed promising results by identifying new EEG and NIRS features (amplitude modulation rate-of-change and polynomial fitting coefficients, respectively) that could be used to complement the functionality of a BCI and increase its usability, especially in terms of temporal demand and perceived performance. Nonetheless, further work with a methodology designed for this purpose will be required to confirm our findings.

5.5 Limitations of this study

In light of the results described above and the conclusion extracted from the literature review (Section 2.2), a number of points must be highlighted that would require methodological adjustments or further investigation. Specifically, we expand here on some details pertaining to the experimental protocol, the preprocessing pipeline and the classification procedure. These elements should be considered in a subsequent iteration of this work or in related endeavours.

5.5.1 Adjustments to the experimental protocol

The recording of neurophysiological signals was done using a cap, provided in three sizes, that could hold both EEG and NIRS probes at the same time. The inion-nasion and ear-to-ear distances were measured for each participant and used to position the cap correctly, so that the Fpz electrode would be located according to the 10-20 International system (see Figure 2.2). However, the cap being available in three different sizes only, it was impossible to perfectly position the rest of the probes to exactly fit the

10-10 system in most participants. Therefore, electrode and optode locations were often not completely accurate or consistent from participant to participant, as each had different head sizes and shapes. This motivated the choice of training and evaluation of classifier performance for each participant individually, instead of across participants. At the same time, this means that spatial analyses of feature importance across participants (Figures 4.3 and 4.6) are not entirely precise. Nonetheless, intra-subject variability being an ubiquitous confounding factor in this type of analysis we assume the inaccuracies introduced by the montage do not significantly impact our interpretation of the results. To alleviate these effects, one should digitize the probe locations relative to head anatomical landmarks at each recording session, and use this information to reproduce a more spatially accurate model of the collected data.

Second, obtaining good signal quality during NIRS data collection proved to be a major challenge. Changes to the setup procedure were applied after collecting data on the first four participants to improve the signal quality and the number of usable channels. Despite these adjustments, NIRS signal quality remained highly dependent on participants' hair type and colour, and on how well the cap was fitted. New cap and sensor technology might make the process more efficient in the coming years.

Third, the three-second countdown occurring before each task trial, although useful to make sure the participants can start performing the mental tasks as soon as they are prompted to, probably resulted in a few subjects starting the task in their head before it even started. This could explain the higher kappas obtained during the last three seconds of the baseline in EEG and feature fusion contexts (see Figures 4.1 and 4.7).

Finally, a potential bias for our study is the presence of task workload-induced drowsiness. It can be seen from the EEG feature importance and ERS/ERS analyses that very strong patterns arise in the α band over the occipital and posterior parietal regions. For instance, this was particularly the case for S05. This parieto-occipital α pattern was especially apparent in the feature importance analysis when a brain-teaser was paired with a more passive imagery task such as mental singing or motor imagery. A reason for this could be the drowsiness induced by tasks that did not require much mental effort, and that might have caused some participants to momentarily close their eyes out of fatigue and boredom. Indeed, it has been shown that strong α power is generated by neural circuits located in the occipital and posterior parietal regions of the brain when one closes one's eyes [94]. Although an experimenter was always present with the participant during the recordings to make sure the instructions were correctly followed, it is possible that some participants closed their eyes periodically without the experimenter noticing. The experimental protocol should thus be adapted to prevent such drowsy periods from occuring. For example, we could have given participants more time to rest by giving them additional breaks during the recording sessions. While we gave them three breaks in each session, this still means data was recorded continuously over periods of 15 minutes, in which participants were instructed to refrain from moving. More breaks probably would have allowed them to fully recover while also limiting their fatigue and boredom. Another improvement would be to reduce the setup time as it often took around one hour to set up the complete multimodal equipment. This part of the experiment is often very boring for a participant, which means that once the actual recording starts they are already tired. Improved NIRS

and EEG equipments (such as new caps, dry electrodes, etc.) could help achieve that goal, as well as prioritizing brain regions of interest to reduce the number of probes used.

5.5.2 Effect of removing bad NIRS channels

For some participants, many NIRS channels had to be removed in the preprocessing stage due to poor signal quality. These rejected channels were often on top of the head, where hair and head shape made it harder to obtain a good contact between the probes and the scalp. This means the analysis of feature importance across subjects can be biased by the fact that not all participants had the same NIRS features available, and that even between a participant's three sessions these might change. Although we agree this does not allow us to get an accurate picture of which brain regions are the most useful to distinguish between two mental tasks, our second best NIRS participant, S04, was actually the participant who had the most NIRS channels rejected (27 on average). Therefore, clearly some discriminative data was found in the remaining channels, and still led to meaningful results. This effect was mitigated in EEG by the ability to interpolate bad channels using neighbouring channels and thus should not be a problem in our analysis.

5.5.3 Classification procedure

Our classification pipeline used a single hyperparameter C for all models (normalized by the number of examples in the training set). This hyperparameter controls the amount of regularization - how forgiving the SVM is to the number of misclassified data points - and at the same time the strength of the l_1 feature selection. This choice was made so the analysis could focus on the evolution of κ across time, the selected features, and the impact of adding NIRS to EEG, while avoiding further division of the data to choose an optimal hyperparameter and then test it on a held-out set.

However, this choice means the reported κ are not optimal. Indeed, each combination of participant, pair of tasks and time window represents a distinct classification problem, and a different hyperparameter for the classification model might lead to a better performance. A solution to this problem is to use nested cross-validation to tune hyperparameters or select between different models (LDA, LR, etc.). For example, non-linear classifiers such as kernel SVMs were shown to work better in certain contexts than linear classifiers [193]. This procedure consists of first dividing the dataset into k_1 partitions of equal size. The first $k_1 - 1$ partitions are called the development set, and the remaining partition, the test set. In a second step, the development set is further divided into k_2 partitions of equal size, and follows a procedure identical to the one described in Section 3.7. The first $k_2 - 1$ partitions are called the training set, and are used to train the classifier; the remaining partition is called the validation set, and is used to assess the performance of the trained classifiers. The hyperparameter configuration or model that scored the highest on average over the k_2 folds is considered the optimal model, and its performance is finally tested in an unbiased manner on the test set. This procedure is repeated k_1 times, and the average test

performance is used as the global classification performance over the given dataset. However, since our analysis focuses on feature importance, it is necessary to obtain the model weights for a unique model of hyperparameter configuration. Therefore, after the unbiased evaluation of classification performance with nested cross-validation, we would still need to apply the same procedure as in Section 3.7 to obtain our final weights.

Second, the dimensionality of the dataset is a major challenge when applying machine learning methods. As exposed in Section 2.1.5, the curse of dimensionality can seriously undermine the training of a classifier. In our case, we relied on three procedures to help alleviate this problem: 1) focusing on standard features such as band powers and chromophore amplitude to constrain the number of features, 2) using SVMs, which are known for their robustness in high dimensionality problems, and 3) using an embedded l_1 feature selection method to select only the most relevant features and at the same time avoid further partitioning of the data (which would be necessary for other feature selection procedures such as filter and wrapper methods [80]). Despite these procedures, the dimensionality of the training set in the feature fusion classification case was of about 86 examples versus 777 features, which is very high.

Other options exist to further alleviate this dimensionality problem. First, dimensionality reduction procedures such as Principal Components Analysis (PCA) and Common Spatial Patterns (CSP, used for EEG), both introduced in Section 2.1.4, could have been used to reduce the number of input features fed into the classifier. These methods find optimal linear combinations of the original features such that a maximal amount of variance is explained by the first component (PCA), or so that the ratio between the variance of the two classes is maximized (CSP). However, these methods would make it hard to analyze the importance of each feature since another layer of weights would be added on top of those learned by the classifier. Also, this means all original features must be kept to compute these components, and thus no computation can be saved or channel removed, two things important to optimize in real-time BCI applications. Second, alternate ways of increasing the amount of data could be considered. Simply more data could be collected to significantly increase the number of data points in the available training set. Collecting multimodal data, however, is a significant investment of time and money. Alternatively, inter-subject classification would allow merging of examples from all nine participants together, effectively increasing the amount of data ninefold. Nevertheless, this type of model is highly affected by inter-subject variability and, in our case, missing channels, rendering the classification problem much harder.

Chapter 6

Conclusion

6.1 Summary of research

In this work, we investigated the use of two non-invasive functional neuroimaging techniques, EEG and NIRS, for the binary classification of seven different mental tasks. We identified optimal mental task pairs across nine participants, for EEG-only, NIRS-only and EEG-NIRS fusion classification schemes, and assessed the impact of a multimodal approach on the classification performance.

Pairs formed of a brain-teaser, i.e., a mental task that requires problem-solving skills (ROT, SUB, WORD), and an imagery task (MI, FACE, SING) consistently yielded the best classification performance for unimodal EEG and NIRS schemes, as well as for a multimodal fusion scheme. In contrast to unimodal performance results at par with those of previous reports, the multimodal approach led to an average increase of 0.03 in peak Cohen's κ when using features extracted from one-second windows (equivalent to a 1.5% accuracy increase in balanced settings). Similarly, a 0.17 increase in κ (8.5% accuracy increase) was obtained when focusing on windows extracted in the last five seconds of the 15-second trials.

The analysis of the trained classification models unveiled interesting spatial patterns of brain activity and the importance of feature subtypes in modality fusion. Particularly, α- and β-related EEG features proved to be the most useful, followed by HbO or HbR amplitude NIRS features, depending on the individual. The occipital and parietal regions yielded the most important EEG features, whereas NIRS features extracted from the prefrontal and frontal regions were the most informative. Subjective task load ratings collected from the participants at each recording session indicated an average preference for the motor imagery, mental rotation and word generation tasks, and confirmed the higher mental load induced by brain-teasers. Finally, EEG and NIRS features that could be used to predict the subjective experience of a BCI user were identified through a correlation analysis, although further investigation will be necessary to validate these findings.

Our proposed feature analysis approach made it possible to delve deeper into the classification results, and shed light on the role of different neurophysiological modalities toward more efficient and flexible BCIs. Moreover, our task pair- and participant-based perspective unveiled subject-specific results that are in line with previous BCI studies.

6.2 Future research directions

To allow for the online implementation of mental task-based EEG-NIRS hBCIs, supplementary steps should be undertaken. First, classification performance can be improved by including additional features extracted from EEG and NIRS, extracting and using information from additional peripheral physiological signals (such as heart rate, respiration, skin conductance response, etc.), and selecting optimal classification models by tuning hyperparameters and comparing different classifiers. The effect of combining different window lengths when extracting features should be further explored to optimize the trade-off between response time and accuracy. Second, classification across subjects, instead of subject-specific classification, should be assessed to establish the feasibility of subject-independent classification, which would be more practical for deployment. Third, the feasibility of our multimodal protocol must be assessed on a more diverse population, including individuals with disabilities who could benefit the most from this technology, to make sure our results are translatable.

Following these steps, an important future research direction should be the implementation of online BCIs based on the optimal mental task pairs we identified. The restrictions induced by a real-time implementation (setup time, computing efficacy, robustness to noise, etc.) will all have to be overcome to yield a truly usable brain-computer interface.

References

[1] Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain–computer interfaces for communication and control. Clinical Neurophysiology. 2002;113(6):767–791.

[2] Wolpaw JR, Wolpaw EW. Brain–computer interfaces: something new under the sun. In: Brain-computer interfaces: principles and practice. Oxford University Press; 2012. p. 3–12.

[3] Van Erp JB, Lotte F, Tangermann M. Brain-computer interfaces: beyond medical applications. Computer. 2012;(4):26–34.

[4] Moghimi S, Kushki A, Marie Guerguerian A, Chau T. A review of EEG-based brain-computer interfaces as access pathways for individuals with severe disabilities. Assistive Technology. 2013; 25(2):99–110.

[5] Coyle S, Ward T, Markham C, McDarby G. On the suitability of near-infrared (NIR) systems for next-generation brain–computer interfaces. Physiological Measurement. 2004;25(4):815.

[6] Hwang HJ, Kim S, Choi S, Im CH. EEG-based brain-computer interfaces: a thorough literature survey. International Journal of Human-Computer Interaction. 2013;29(12):814–826.

[7] Leuthardt EC, Schalk G, Wolpaw JR, Ojemann JG, Moran DW. A brain–computer interface using electrocorticographic signals in humans. Journal of Neural Engineering. 2004;1(2):63.

[8] Mellinger J, Schalk G, Braun C, Preissl H, Rosenstiel W, Birbaumer N, Kübler A. An MEG-based brain–computer interface (BCI). Neuroimage. 2007;36(3):581–593.

[9] Sitaram R, Caria A, Birbaumer N. Hemodynamic brain–computer interfaces for communication and rehabilitation. Neural Networks. 2009;22(9):1320–1328.

[10] Sitaram R, Caria A, Veit R, Gaber T, Rota G, Kuebler A, Birbaumer N. fMRI brain-computer interface: a tool for neuroscientific research and treatment. Computational Intelligence and Neuroscience. 2007;2007:10 pages.

[11] Naseer N, Hong KS. fNIRS-based brain-computer interfaces: a review. Frontiers in Human Neuroscience. 2015;9.

[12] Pfurtscheller G, Allison BZ, Bauernfeind G, Brunner C, Solis Escalante T, Scherer R, Zander TO, Müller-Putz G, Neuper C, Birbaumer N. The hybrid BCI. Frontiers in Neuroscience. 2010;4(3):3.

[13] Müller-Putz GR, Leeb R, Millán JR, Horki P, Kreilinger A, Bauernfeind G, Allison BZ, Brunner C, Scherer R. Principles of hybrid brain–computer interfaces. In: Towards practical brain-computer interfaces. Springer; 2013. p. 355–373.

[14] Banville H, Falk TH. Recent advances and open challenges in hybrid brain-computer interfacing: A technological review. Brain-Computer Interfaces. 2016;Submitted for publication.

[15] Kubler A, Mushahwar V, Hochberg LR, Donoghue JP. BCI meeting 2005-workshop on clinical issues and applications. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2006;14(2):131.

[16] Brouwer AM, Zander TO, Van Erp JBF, Korteling H, Bronkhorst AW. Using neurophysiological signals that reflect cognitive or affective state: Six recommendations to avoid common pitfalls. Frontiers in Neuroscience. 2015;9(136).

[17] Fazli S, Mehnert J, Steinbrink J, Curio G, Villringer A, Müller KR, Blankertz B. Enhanced performance by a hybrid NIRS-EEG brain computer interface. Neuroimage. 2012;59(1):519–529.

[18] Khan MJ, Hong MJ, Hong KS. Decoding of four movement directions using hybrid NIRS-EEG brain-computer interface. Frontiers in Human Neuroscience. 2014;8.

[19] Koo B, Lee HG, Nam Y, Kang H, Koh CS, Shin HC, Choi S. A hybrid NIRS-EEG system for self-paced brain computer interface with online motor imagery. Journal of Neuroscience Methods. 2015;244:26–32.

[20] Pasqualotto E, Federici S, Belardinelli MO. Toward functioning and usable brain-computer interfaces (BCIs): a literature review. Disability and Rehabilitation: Assistive Technology. 2012; 7(2):89–103.

[21] Mühl C, Gürkök H, Bos DPO, Thurlings ME, Scherffig L, Duvinage M, Elbakyan AA, Kang S, Poel M, Heylen D. Bacteria hunt: evaluating multi-paradigm BCI interaction. Journal on Multimodal User Interfaces. 2010;4(1):11–25.

[22] Kreilinger A, Kaiser V, Breitwieser C, Williamson J, Neuper C, Müller-Putz GR. Switching between manual control and brain-computer interface using long term and short term quality measures. Frontiers in Neuroscience. 2011;5:147.

[23] Mühl C, Allison B, Nijholt A, Chanel G. A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges. Brain-Computer Interfaces. 2014;1(2):66–84.

[24] Bashashati A, Fatourechi M, Ward RK, Birch GE. A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals. Journal of Neural Engineering. 2007; 4(2):R32.

[25] Müller-Putz GR, Breitwieser C, Cincotti F, Leeb R, Schreuder M, Leotta F, Tavella M, Bianchi L, Kreilinger A, Ramsay A, et al. Tools for brain-computer interaction: a general concept for a hybrid BCI. Frontiers in Neuroinformatics. 2011;5:30.

[26] Mason SG, Birch GE. A general framework for brain-computer interface design. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2003;11(1):70–85.

[27] Malmivuo J, Plonsey R. Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields. Oxford University Press; 1995.

[28] Jackson AF, Bolger DJ. The neurophysiological bases of EEG and EEG measurement: A review for the rest of us. Psychophysiology. 2014;51(11):1061–1071.

[29] Handy T. Event-related potentials: A methods handbook. MIT Press. A Bradford book; 2005. Chapter 11; Available from: `https://books.google.ca/books?id=0QyZEfgEzRUC`.

[30] Friedrich EV, Scherer R, Neuper C. Stability of event-related (de-) synchronization during brain–computer interface-relevant mental tasks. Clinical Neurophysiology. 2013;124(1):61–69.

[31] Basar E, Basar-Eroglu C, Karakas S, Schurmann M. Gamma, alpha, delta, and theta oscillations govern cognitive processes. International Journal of Psychophysiology. 2001;39(2–3):241 – 248.

[32] Ward TE. Hybrid optical–electrical brain computer interfaces, practices and possibilities. In: Towards practical brain-computer interfaces. Springer; 2013. p. 17–40.

[33] Scholkmann F, Kleiser S, Metz AJ, Zimmermann R, Pavia JM, Wolf U, Wolf M. A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. Neuroimage. 2014;85:6–27.

[34] Ferrari M, Quaresima V. A brief review on the history of human functional near-infrared spectroscopy (fnirs) development and fields of application. Neuroimage. 2012;63(2):921–935.

[35] Vialatte FB, Maurice M, Dauwels J, Cichocki A. Steady-state visually evoked potentials: focus on essential paradigms and future perspectives. Progress in Neurobiology. 2010;90(4):418–438.

[36] Friedrich EV, Scherer R, Neuper C. The effect of distinct mental strategies on classification performance for brain–computer interfaces. International Journal of Psychophysiology. 2012;84(1):86–94.

[37] Van Gerven M, Bahramisharif A, Heskes T, Jensen O. Selecting features for BCI control based on a covert spatial attention paradigm. Neural Networks. 2009;22(9):1271–1277.

[38] Dockstader C, Cheyne D, Tannock R. Cortical dynamics of selective attention to somatosensory events. Neuroimage. 2010;49(2):1777–1785.

[39] Fatourechi M, Bashashati A, Ward RK, Birch GE. EMG and EOG artifacts in brain computer interface systems: A survey. Clinical Neurophysiology. 2007;118(3):480–494.

[40] Urigüen JA, Garcia-Zapirain B. EEG artifact removal—state-of-the-art and guidelines. Journal of Neural Engineering. 2015;12(3):031001.

[41] Cooper RJ, Selb J, Gagnon L, Phillip D, Schytz HW, Iversen HK, Ashina M, Boas DA. A systematic comparison of motion artifact correction techniques for functional near-infrared spectroscopy. Frontiers in Neuroscience. 2012;6.

[42] Lotte F, Congedo M, Lécuyer A, Lamarche F, Arnaldi B, et al. A review of classification algorithms for EEG-based brain–computer interfaces. Journal of Neural Engineering. 2007;4:R1–R13.

[43] Cortes C, Vapnik V. Support-vector networks. Machine learning. 1995;20(3):273–297.

[44] Gürkök H, Nijholt A. Brain–computer interfaces for multimodal interaction: a survey and principles. International Journal of Human-Computer Interaction. 2012;28(5):292–307.

[45] Obermaier B, Neuper C, Guger C, Pfurtscheller G. Information transfer rate in a five-classes brain-computer interface. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2001; 9(3):283–288.

[46] Curran E, Sykacek P, Stokes M, Roberts SJ, Penny W, Johnsrude I, Owen AM. Cognitive tasks for driving a brain-computer interfacing system: a pilot study. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2004;12(1):48–54.

[47] Sepulveda F, Dyson M, Gan JQ, Tsui CL. A comparison of mental task combinations for asynchronous EEG-based BCIs. In: Engineering in Medicine and Biology Society, 2007. 29th Annual International Conference of the IEEE; 2007. p. 5055–5058.

[48] Faradji F, Ward RK, Birch GE. A brain-computer interface based on mental tasks with a zero false activation rate. In: 4th International IEEE/EMBS Conference on Neural Engineering, 2009. NER'09. IEEE; 2009. p. 355–358.

[49] Dobrea MC, Dobrea DM. The selection of proper discriminative cognitive tasks - a necessary prerequisite in high-quality BCI applications. In: Applied Sciences in Biomedical and Communication Technologies, 2009. ISABEL 2009. 2nd International Symposium on. IEEE; 2009. p. 1–6.

[50] Friedrich EV, Scherer R, Sonnleitner K, Neuper C. Impact of auditory distraction on user performance in a brain–computer interface driven by different mental tasks. Clinical Neurophysiology. 2011;122(10):2003–2009.

[51] Scherer R, Faller J, Balderas D, Friedrich EV, Pröll M, Allison B, Müller-Putz G. Brain–computer interfacing: more than the sum of its parts. Soft Computing. 2013;17(2):317–331.

[52] Friedrich EV, Scherer R, Neuper C. Long-term evaluation of a 4-class imagery-based brain–computer interface. Clinical Neurophysiology. 2013;124(5):916–927.

[53] Friedrich EVC, Neuper C, Scherer R. Whatever works: A systematic user-centered training protocol to optimize brain-computer interfacing individually. PLoS ONE. 2013;8(9):e76214.

[54] Hoshi Y, Tamura M. Near-infrared optical detection of sequential brain activation in the prefrontal cortex during mental tasks. NeuroImage. 1997;5(4):292–297.

[55] Abibullaev B, An J, Moon JI. Neural network classification of brain hemodynamic responses from four mental tasks. International Journal of Optomechatronics. 2011;5(4):340–359.

[56] Power S, Kushki A, Chau T. Toward a system-paced NIRS-BCI: differentiating prefrontal activity due to mental artithmetic and music imagery from the no-control state. Journal of Neural Engineering. 2011;8(066004):14.

[57] Falk TH, Guirgis M, Power S, Chau TT. Taking NIRS-BCIs outside the lab: towards achieving robustness against environment noise. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2011;19(2):136–146.

[58] Power SD, Kushki A, Chau T. Intersession consistency of single-trial classification of the prefrontal response to mental arithmetic and the no-control state by NIRS. PloS one. 2012;7(7):e37791.

[59] Herff C, Heger D, Putze F, Hennrich J, Fortmann O, Schultz T. Classification of mental tasks in the prefrontal cortex using fNIRS. In: Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE; 2013. p. 2160–2163.

[60] Schudlo LC, Power SD, Chau T. Dynamic topographical pattern classification of multichannel prefrontal NIRS signals. Journal of Neural Engineering. 2013;10(4):046018.

[61] Hwang HJ, Lim JH, Kim DW, Im CH. Evaluation of various mental task combinations for near-infrared spectroscopy-based brain-computer interfaces. Journal of Biomedical Optics. 2014; 19(7):077005–077005.

[62] Schudlo LC, Chau T. Single-trial classification of near-infrared spectroscopy signals arising from multiple cortical regions. Behavioural brain research. 2015;290:131–142.

[63] Myrden A, Kushki A, Sejdić E, Chau T. Towards increased data transmission rate for a three-class metabolic brain–computer interface based on transcranial doppler ultrasound. Neuroscience letters. 2012;528(2):99–103.

[64] Aleem I, Chau T. Towards a hemodynamic BCI using transcranial doppler without user-specific training data. Journal of neural engineering. 2013;10(1):016005.

[65] Faress A, Chau T. Towards a multimodal brain-computer interface: combining fNIRS and fTCD measurements to enable higher classification accuracy. Neuroimage. 2013;77(0):186–194.

[66] Nawa NE, Ando H. Classification of self-driven mental tasks from whole-brain activity patterns. PLoS ONE. 2014;9(5):e97296.

[67] Zacks JM. Neuroimaging studies of mental rotation: a meta-analysis and review. Journal of Cognitive Neuroscience. 2008;20(1):1–19.

[68] Pietsch S, Jansen P. Different mental rotation performance in students of music, sport and education. Learning and Individual Differences. 2012;22(1):159–163.

[69] Hart SG, Staveland LE. Development of NASA-TLX (task load index): Results of empirical and theoretical research. Advances in Psychology. 1988;52:139–183.

[70] Cegarra J, Morgado N. Étude des propriétés de la version francophone du NASA TLX. In: Communication présentée à la cinquième édition du colloque de psychologie ergonomique (Epique); 2009.

[71] Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods. 2004;134(1):9–21.

[72] Mognon A, Jovicich J, Bruzzone L, Buiatti M. ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. Psychophysiology. 2011;48(2):229–240.

[73] Kalcher J, Pfurtscheller G. Discrimination between phase-locked and non-phase-locked event-related EEG activity. Electroencephalography and Clinical Neurophysiology. 1995;94(5):381–384.

[74] Falk TH, Fraga FJ, Trambaiolli L, Anghinah R. EEG amplitude modulation analysis for semi-automated diagnosis of alzheimer's disease. EURASIP Journal on Advances in Signal Processing. 2012;2012(1):1–9.

[75] Cassani R, Falk TH, Fraga FJ, Kanda PA, Anghinah R. The effects of automated artifact removal algorithms on electroencephalography-based alzheimer's disease diagnosis. Frontiers in Aging Neuroscience. 2014;6.

[76] Weiss S, Mueller HM. The contribution of EEG coherence to the investigation of language. Brain and language. 2003;85(2):325–343.

[77] Seemuller A, Muller E, Rosler F. EEG-power and -coherence changes in a unimodal and a cross-modal working memory task with visual and kinesthetic stimuli. International Journal of Psychophysiology. 2012;83(1):87 – 95.

[78] Koenig T, Lehmann D, Saito N, Kuginuki T, Kinoshita T, Koukkou M. Decreased functional connectivity of EEG theta-frequency activity in first-episode, neuroleptic-naıve patients with schizophrenia: preliminary results. Schizophrenia Research. 2001;50(1–2):55 – 60.

[79] Huppert TJ, Diamond SG, Franceschini MA, Boas DA. HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. Applied Optics. 2009;48(10):D280–D298.

[80] Guyon I, Elisseeff A. An introduction to variable and feature selection. The Journal of Machine Learning Research. 2003;3:1157–1182.

[81] Ng AY. Feature selection, l1 vs. l2 regularization, and rotational invariance. In: Proceedings of the twenty-first international conference on Machine learning. ACM; 2004. p. 78.

[82] Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Machine learning. 2002;46(1-3):389–422.

[83] Price CJ. The anatomy of language: a review of 100 fMRI studies published in 2009. Annals of the New York Academy of Sciences. 2010;1191(1):62–88.

[84] Fehr T, Code C, Herrmann M. Common brain regions underlying different arithmetic operations as revealed by conjunct fMRI–BOLD activation. Brain research. 2007;1172:93–102.

[85] Curtis CE, D'Esposito M. Persistent activity in the prefrontal cortex during working memory. Trends in cognitive sciences. 2003;7(9):415–423.

[86] Kleber B, Birbaumer N, Veit R, Trevorrow T, Lotze M. Overt and imagined singing of an italian aria. Neuroimage. 2007;36(3):889–900.

[87] Gunji A, Ishii R, Chau W, Kakigi R, Pantev C. Rhythmic brain activities related to singing in humans. Neuroimage. 2007;34(1):426–434.

[88] Ino T, Inoue Y, Kage M, Hirose S, Kimura T, Fukuyama H. Mental navigation in humans is processed in the anterior bank of the parieto-occipital sulcus. Neuroscience Letters. 2002;322(3):182–186.

[89] Pfurtscheller G, Neuper C. Motor imagery activates primary sensorimotor area in humans. Neuroscience Letters. 1997;239(2–3):65 – 68.

[90] Pfurtscheller G, Brunner C, Schlögl A, da Silva FL. Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks. NeuroImage. 2006;31(1):153 – 159.

[91] Ishai A, Haxby JV, Ungerleider LG. Visual imagery of famous faces: effects of memory and attention revealed by fMRI. Neuroimage. 2002;17(4):1729–1741.

[92] Pessoa L, Gutierrez E, Bandettini PA, Ungerleider LG. Neural correlates of visual working memory: fMRI amplitude predicts task performance. Neuron. 2002;35(5):975 – 987.

[93] Yu R, Mobbs D, Seymour B, Rowe JB, Calder AJ. The neural signature of escalating frustration in humans. Cortex. 2014;54:165 – 178.

[94] Kalauzi A, Vuckovic A, Bojić T. EEG alpha phase shifts during transition from wakefulness to drowsiness. International Journal of Psychophysiology. 2012;86(3):195 – 205.

[95] Banville H, Gupta R, Falk TH. Nirs-eeg multimodal evaluation of mental tasks for hybrid brain-computer interfacing. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2016; In preparation.

[96] Nicolas-Alonso LF, Gomez-Gil J. Brain computer interfaces, a review. Sensors. 2012;12(2):1211–1279.

[97] Hämäläinen M, Hari R, Ilmoniemi RJ, Knuutila J, Lounasmaa OV. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. Reviews of Modern Physics. 1993;65(2):413.

[98] Glover GH. Overview of functional magnetic resonance imaging. Neurosurgery Clinics of North America. 2011;22(2):133–139.

[99] Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiological Measurement. 2007;28(3):R1.

[100] Blain S, Mihailidis A, Chau T. Assessing the potential of electrodermal activity as an alternative access pathway. Medical Engineering & Physics. 2008;30(4):498–505.

[101] Lodish H ZS Berk A. Molecular cell biology. 4th ed. W. H. Freeman; 2000. Chapter Section 21.1, Overview of Neuron Structure and Function; an optional note.

[102] Purves D, Fitzpatrick D, Katz L, Lamantia A, McNamara J, Williams S, Augustine G. Neuroscience. Sinauer Associates; 2001; Available from: `https://books.google.ca/books?id=F4pTPwAACAAJ`.

[103] Wassenaar E, Van den Brand J. Reliability of near-infrared spectroscopy in people with dark skin pigmentation. Journal of Clinical Monitoring and Computing. 2005;19(3):195–199.

[104] Schalk G, Wolpaw JR, McFarland DJ, Pfurtscheller G. EEG-based communication: presence of an error potential. Clinical Neurophysiology. 2000;111(12):2138–2144.

[105] Blankertz B, Schäfer C, Dornhege G, Curio G. Single trial detection of EEG error potentials: A tool for increasing BCI transmission rates. In: Artificial neural networks—ICANN 2002. Springer; 2002. p. 1137–1143.

[106] Dal Seno B, Matteucci M, Mainardi L. Online detection of P300 and error potentials in a BCI speller. Computational Intelligence and Neuroscience. 2010;2010:11.

[107] Schlögl A, Lee F, Bischof H, Pfurtscheller G. Characterization of four-class motor imagery EEG data for the BCI-competition 2005. Journal of Neural Engineering. 2005;2(4):L14.

[108] Birbaumer N. Slow cortical potentials: plasticity, operant control, and behavioral effects. The Neuroscientist. 1999;5(2):74–78.

[109] Cabeza R, Nyberg L. Imaging cognition II: an empirical review of 275 PET and fMRI studies. Journal of Cognitive Neuroscience. 2000;12(1):1–47.

[110] Aguirre GK, Zarahn E, D'esposito M. The variability of human, BOLD hemodynamic responses. Neuroimage. 1998;8(4):360–369.

[111] Coyle SM, Ward TE, Markham CM. Brain–computer interface using a simplified functional near-infrared spectroscopy system. Journal of Neural Engineering. 2007;4(3):219–226.

[112] Nishimoto S, Vu A, Naselaris T, Benjamini Y, Yu B, Gallant J. Reconstructing visual experiences from brain activity evoked by natural movies. Current Biology. 2011;21(19):1641 – 1646.

[113] Abibullaev B, An J. Classification of frontal cortex haemodynamic responses during cognitive tasks using wavelet transforms and machine learning algorithms. Medical Engineering & Physics. 2012; 34(10):1394–1410.

[114] Carvalhaes C, de Barros JA. The surface laplacian technique in EEG: Theory and methods. International Journal of Psychophysiology. 2015;97(3):174 – 188.

[115] Wang Y, Gao S, Gao X. Common spatial pattern method for channel selelction in motor imagery based brain-computer interface. In: Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference; 2006. p. 5392–5395.

[116] Friedman JH. On bias, variance, 0/1-loss, and the curse-of-dimensionality. Data Mining and Knowledge Discovery. 1997;1(1):55–77.

[117] Guyon I. SVM application list. 2006; accessed: 2015-11-13; Available from: `http://www.clopinet.com/SVM.applications.html`.

[118] Commons W. SVM max sep hyperplane with margin. 2011; file: LambdaPlaques.jpg; Available from: `https://commons.wikimedia.org/wiki/File:Svm_max_sep_hyperplane_with_margin.png`.

[119] Zander TO, Kothe C, Welke S, Roetting M. Enhancing human-machine systems with secondary input from passive brain-computer interfaces. In: Proceedings of the 4th International BCI Workshop & Training Course; 2008. p. 144–149.

[120] Kleber B, Birbaumer N. Direct brain communication: neuroelectric and metabolic approaches at tübingen. Cognitive Processing. 2005;6(1):65–74.

[121] Lesenfants D, Habbal D, Lugo Z, Lebeau M, Horki P, Amico E, Pokorny C, Gómez F, Soddu A, Müller-Putz G, et al. An independent SSVEP-based brain–computer interface in locked-in syndrome. Journal of Neural Engineering. 2014;11(3):035002.

[122] Treder MS, Schmidt NM, Blankertz B. Gaze-independent brain–computer interfaces based on covert attention and feature attention. Journal of Neural Engineering. 2011;8(6):066003.

[123] Thompson DE, Quitadamo LR, Mainardi L, Gao S, Kindermans PJ, Simeral JD, Fazel-Rezai R, Matteucci M, Falk TH, Bianchi L, et al. Performance measurement for brain–computer or brain–machine interfaces: a tutorial. Journal of Neural Engineering. 2014;11(3):035001.

[124] Nijboer F, Clausen J, Allison BZ, Haselager P. The Asilomar survey: stakeholders' opinions on ethical issues related to brain-computer interfacing. Neuroethics. 2013;6(3):541–578.

[125] Wolpaw JR, Birbaumer N, Heetderks WJ, McFarland DJ, Peckham PH, Schalk G, Donchin E, Quatrano LA, Robinson CJ, Vaughan TM, et al. Brain-computer interface technology: a review of the first international meeting. IEEE Transactions on Rehabilitation Engineering. 2000;8(2):164–173.

[126] Zander TO, Kothe C, Jatzev S, Gaertner M. Enhancing human-computer interaction with input from active and passive brain-computer interfaces. In: Brain-computer interfaces. Springer; 2010. p. 181–199.

[127] Wolpaw JR. The BCI endeavor and the mission of this new journal. Brain-Computer Interfaces. 2014;1(1):2–4.

[128] Allison BZ, Brunner C, Kaiser V, Müller-Putz GR, Neuper C, Pfurtscheller G. Toward a hybrid brain-computer interface based on imagined movement and visual attention. Journal of Neural Engineering. 2010;7(2):026007.

[129] Leeb R, Sagha H, Chavarriaga R, del R Millán J. A hybrid brain-computer interface based on the fusion of electroencephalographic and electromyographic activities. Journal of Neural Engineering. 2011;8(2):025011.

[130] Maye A, Zhang D, Wang Y, Gao S, Engel AK. Multimodal brain-computer interfaces. Tsinghua Science and Technology. 2011;16(2):133–139.

[131] Brunner C, Allison BZ, Krusienski DJ, Kaiser V, Müller-Putz GR, Pfurtscheller G, Neuper C. Improved signal processing approaches in an offline simulation of a hybrid brain-computer interface. Journal of Neuroscience Methods. 2010;188(1):165–173.

[132] Horki P, Solis-Escalante T, Neuper C, Müller-Putz G. Combined motor imagery and SSVEP based BCI control of a 2 DoF artificial upper limb. Medical & Biological Engineering & Computing. 2011;49(5):567–577.

[133] Brunner C, Allison BZ, Altstatter C, Neuper C. A comparison of three brain-computer interfaces based on event-related desynchronization, steady state visual evoked potentials, or a hybrid approach using both signals. Journal of Neural Engineering. 2011;8(2):025010.

[134] Allison BZ, Leeb R, Brunner C, Müller-Putz GR, Bauernfeind G, Kelly JW, Neuper C. Toward smarter BCIs: extending BCIs through hybridization and intelligent control. Journal of Neural Engineering. 2012;9(1):013001.

[135] Cao T, Wan F, Wong CM, da Cruz JN, Hu Y. Objective evaluation of fatigue by EEG spectral analysis in steady-state visual evoked potential-based brain-computer interfaces. Biomedical Engineering Online. 2014;13(1):28.

[136] Li J, Ji H, Cao L, Zang D, Gu R, Xia B, Wu Q. Evaluation and application of a hybrid brain computer interface for real wheelchair parallel control with multi-degree of freedom. International Journal of Neural Systems. 2014;24(04):1450014.

[137] Li Y, Long J, Yu T, Yu Z, Wang C, Zhang H, Guan C. An EEG-based BCI system for 2D cursor control by combining mu/beta rhythm and P300 potential. IEEE Transactions on Biomedical Engineering. 2010;57(10):2495–2505.

[138] Su Y, Qi Y, Luo JX, Wu B, Yang F, Li Y, Zhuang YT, Zheng XX, Chen WD. A hybrid brain-computer interface control strategy in a virtual environment. Journal of Zhejiang University: Science C. 2011;12(5):351–361.

[139] Yu T, Li Y, Long J, Gu Z. Surfing the internet with a BCI mouse. Journal of Neural Engineering. 2012;9(3):036012.

[140] Long JY, Li YQ, Yu TY, Gu ZH. Target selection with hybrid feature for BCI-based 2D cursor control. IEEE Transactions on Biomedical Engineering. 2012;59(1):132–140.

[141] Long JY, Li YQ, Wang HT, Yu TY, Pan JH, Li F. A hybrid brain computer interface to control the direction and speed of a simulated or real wheelchair. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2012;20(5):720–729.

[142] Yu TY, Li YQ, Long JY, Li F. A hybrid brain-computer interface-based mail client. Computational and Mathematical Methods in Medicine. 2013;2013:9.

[143] Lorenz R, Pascual J, Blankertz B, Vidaurre C. Towards a holistic assessment of the user experience with hybrid BCIs. Journal of Neural Engineering. 2014;11(3):035007.

[144] Yao L, Meng J, Zhang D, Sheng X, Zhu X. Combining motor imagery with selective sensation towards a hybrid-modality BCI. IEEE Transactions on Biomedical Engineering. 2014;61(8):2304–2312.

[145] Combaz A, Chumerin N, Manyakov NV, Robben A, Suykens JA, Van Hulle MM. Towards the detection of error-related potentials and its integration in the context of a P300 speller brain–computer interface. Neurocomputing. 2012;80:73–82.

[146] Spüler M, Bensch M, Kleih S, Rosenstiel W, Bogdan M, Kübler A. Online use of error-related potentials in healthy users and people with severe motor impairment increases performance of a P300-BCI. Clinical Neurophysiology. 2012;123(7):1328–1337.

[147] Rutkowski TM, Mori H. Tactile and bone-conduction auditory brain computer interface for vision and hearing impaired users. Journal of Neuroscience Methods. 2014;S0165-0270(14):00126–5.

[148] Panicker RC, Puthusserypady S, Sun Y. An asynchronous P300 BCI with SSVEP-based control state detection. IEEE Transactions on Biomedical Engineering. 2011;58(6):1781–1788.

[149] Xu M, Qi H, Wan B, Yin T, Liu Z, Ming D. A hybrid BCI speller paradigm combining P300 potential and the SSVEP blocking feature. Journal of Neural Engineering. 2013;10(2):026001.

[150] Yin E, Zhou Z, Jiang J, Chen F, Liu Y, Hu D. A novel hybrid BCI speller based on the incorporation of SSVEP into the P300 paradigm. Journal of Neural Engineering. 2013;10(2):026012.

[151] Li YQ, Pan JH, Wang F, Yu ZL. A hybrid BCI system combining P300 and SSVEP and its application to wheelchair control. IEEE Transactions on Biomedical Engineering. 2013;60(11):3156–3166.

[152] Pan J, Xie Q, He Y, Wang F, Di H, Laureys S, Yu R, Li Y. Detecting awareness in patients with disorders of consciousness using a hybrid brain–computer interface. Journal of Neural Engineering. 2014;11(5):056007.

[153] Xu M, Chen L, Zhang L, Qi H, Ma L, Tang J, Wan B, Ming D. A visual parallel-BCI speller based on the time–frequency coding strategy. Journal of Neural Engineering. 2014;11(2):026014.

[154] Yin E, Zhou Z, Jiang J, Chen F, Liu Y, Hu D. A speedy hybrid BCI spelling approach combining P300 and SSVEP. IEEE Transactions on Biomedical Engineering. 2014;61(2):473–483.

[155] Choi B, Jo S. A low-cost EEG system-based hybrid brain-computer interface for humanoid robot navigation and recognition. PLoS ONE. 2013;8(9):e74583.

[156] Postelnicu CC, Talaba D. P300-based brain-neuronal computer interaction for spelling applications. IEEE Transactions on Biomedical Engineering. 2013;60(2):534–543.

[157] Jiang J, Zhou Z, Yin E, Yu Y, Hu D. Hybrid brain-computer interface (BCI) based on the EEG and EOG signals. Bio-medical Materials and Engineering. 2014;24(6):2919–2925.

[158] Wang H, Li Y, Long J, Yu T, Gu Z. An asynchronous wheelchair control by hybrid EEG–EOG brain–computer interface. Cognitive Neurodynamics. 2014;8(5):399–409.

[159] Perdikis S, Leeb R, Williamson J, Ramsay A, Tavella M, Desideri L, Hoogerwerf EJ, Al-Khodairy A, Murray-Smith R, d R Millán J. Clinical evaluation of braintree, a motor imagery hybrid BCI speller. Journal of Neural Engineering. 2014;11(3):036003.

[160] Severini G, Conforto S, Schmid M, D'Alessio T. A multivariate auto-regressive method to estimate cortico-muscular coherence for the detection of movement intent. Applied Bionics and Biomechanics. 2012;9(2):135–143.

[161] Gallego JA, Ibanez J, Dideriksen JL, Serrano JI, del Castillo MD, Farina D, Rocon E. A multimodal human-robot interface to drive a neuroprosthesis for tremor management. IEEE Transactions on Systems Man and Cybernetics Part C-Applications and Reviews. 2012;42(6):1159–1168.

[162] Yong X, Fatourechi M, Ward RK, Birch GE. The design of a point-and-click system by integrating a self-paced brain-computer interface with an eye-tracker. IEEE Journal on Emerging and Selected Topics in Circuits and Systems. 2011;1(4):590–602.

[163] Yong X, Fatourechi M, Ward RK, Birch GE. Automatic artefact removal in a self-paced hybrid brain- computer interface system. Journal of NeuroEngineering and Rehabilitation. 2012;9(1):1–20.

[164] Qian M, Aguilar M, Zachery KN, Privitera C, Klein S, Carney T, Nolte LW. Decision-level fusion of EEG and pupil features for single-trial visual detection analysis. IEEE Transactions on Biomedical Engineering. 2009;56(7):1929–1937.

[165] Hyung Kim B, Kim M, Joevaluated S. Quadcopter flight control using a low-cost hybrid interface with EEG-based classification and eye tracking. Computers in Biology and Medicine. 2014; 51(0):82–92.

[166] Leeb R, Lancelle M, Kaiser V, Fellner DW, Pfurtscheller G. Thinking penguin: multimodal brain-computer interface control of a VR game. IEEE Transactions on Computational Intelligence and AI in Games. 2013;5(2):117–128.

[167] Rohm M, Schneiders M, Müller C, Kreilinger A, Kaiser V, Müller-Putz GR, Rupp R. Hybrid brain-computer interfaces and hybrid neuroprostheses for restoration of upper limb functions in individuals with high-level spinal cord injury. Artificial Intelligence in Medicine. 2013;59(2):133–142.

[168] Tomita Y, Vialatte FB, Dreyfus G, Mitsukura Y, Bakardjian H, Cichocki A. Bimodal BCI using simultaneously NIRS and EEG. IEEE Transactions on Biomedical Engineering. 2014;61(4):1274–1284.

[169] Zimmermann R, Marchal-Crespo L, Edelmann J, Lambercy O, Fluet MC, Riener R, Wolf M, Gassert R, et al. Detection of motor execution using a hybrid fNIRS-biosignal BCI: a feasibility study. Journal Neuroengineering and Rehabilitation. 2013;10(4).

[170] Rutkowski TM, Tanaka T, Cichocki A, Erickson D, Cao J, Mandic DP. Interactive component extraction from fEEG, fNIRS and peripheral biosignals for affective brain–machine interfacing paradigms. Computers in Human Behavior. 2011;27(5):1512–1518.

[171] Mason SG, Bashashati A, Fatourechi M, Navarro KF, Birch GE. A comprehensive survey of brain interface technology designs. Annals of Biomedical Engineering. 2007;35(2):137–169.

[172] Sekhon LHS, Fehlings MG. Epidemiology, demographics, and pathophysiology of acute spinal cord injury. Spine. 2001;26(24S):S2–S12.

[173] Bernardi L, Wdowczyk-Szulc J, Valenti C, Castoldi S, Passino C, Spadacini G, Sleight P. Effects of controlled breathing, mental activity and mental stress with or without verbalization on heart rate variability. Journal of the American College of Cardiology. 2000;35(6):1462–1469.

[174] Mak JN, Wolpaw JR. Clinical applications of brain-computer interfaces: current state and future prospects. IEEE Reviews in Biomedical Engineering. 2009;2:187–199.

[175] Blankertz B, Tangermann M, Vidaurre C, Fazli S, Sannelli C, Haufe S, Maeder C, Ramsey L, Sturm I, Curio G, et al. The berlin brain–computer interface: non-medical uses of BCI technology. Frontiers in Neuroscience. 2010;4.

[176] Millán JR, Rupp R, Müller-Putz GR, Murray-Smith R, Giugliemma C, Tangermann M, Vidaurre C, Cincotti F, Kübler A, Leeb R, et al. Combining brain–computer interfaces and assistive technologies: state-of-the-art and challenges. Frontiers in Neuroscience. 2010;4(161).

[177] Machado S, Almada LF, Annavarapu RN. Progress and prospects in EEG-based brain-computer interface: clinical applications in neurorehabilitation. Journal of Rehabilitation Robotics. 2013; 1(1):28–41.

[178] Rivet B, Souloumiac A, Attina V, Gibert G. xDAWN algorithm to enhance evoked potentials: application to brain–computer interface. IEEE Transactions on Biomedical Engineering. 2009; 56(8):2035–2043.

[179] Mackert BM, Wübbeler G, Leistner S, Uludag K, Obrig H, Villringer A, Trahms L, Curio G. Neurovascular coupling analyzed non-invasively in the human brain. Neuroreport. 2004;15(1):63–66.

[180] Kroupi E, Vesin JM, Ebrahimi T. Implicit affective profiling of subjects based on physiological data coupling. Brain-Computer Interfaces. 2014;1(2):85–98.

[181] Thomas E, Dyson M, Clerc M. An analysis of performance evaluation for motor-imagery based BCI. Journal of Neural Engineering. 2013;10(3):031001.

[182] Jerritta S, Murugappan M, Nagarajan R, Wan K. Physiological signals based human emotion recognition: a review. In: Signal Processing and its Applications (CSPA), 2011 IEEE 7th International Colloquium on; March; 2011. p. 410–415.

[183] Daly I, Scherer R, Billinger M, Müller-Putz G. FORCe: fully online and automated artifact removal for brain-computer interfacing. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2014;PP(99):1–1.

[184] Sajda P, Gerson A, Müller K, Blankertz B, Parra L. A data analysis competition to evaluate machine learning algorithms for use in brain-computer interfaces. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2003;11(2):184–185.

[185] Blankertz B, Müller K, Curio G, Vaughan TM, Schalk G, Wolpaw JR, Schlögl A, Neuper C, Pfurtscheller G, Hinterberger T, et al. The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials. IEEE Transactions on Biomedical Engineering. 2004;51(6):1044–1051.

[186] Blankertz B, Müller K, Krusienski DJ, Schalk G, Wolpaw JR, Schlögl A, Pfurtscheller G, Millán JR, Schroder M, Birbaumer N. The BCI competition III: Validating alternative approaches to actual BCI problems. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2006; 14(2):153–159.

[187] Koelstra S, Mühl C, Soleymani M, Lee JS, Yazdani A, Ebrahimi T, Pun T, Nijholt A, Patras I. DEAP: A database for emotion analysis; using physiological signals. IEEE Transactions on Affective Computing. 2012;3(1):18–31.

[188] Soleymani M, Lichtenauer J, Pun T, Pantic M. A multimodal database for affect recognition and implicit tagging. IEEE Transactions on Affective Computing. 2012;3(1):42–55.

[189] Gupta R, Banville H, Falk TH. Physyqx: A database for physiological evaluation of synthesized speech quality-of-experience. In: IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA); 2015.

[190] Pfurtscheller G, Da Silva FL. Event-related EEG/MEG synchronization and desynchronization: basic principles. Clinical Neurophysiology. 1999;110(11):1842–1857.

[191] Braun ML, Buhmann JM, Müller KR. On relevant dimensions in kernel feature spaces. The Journal of Machine Learning Research. 2008;9:1875–1908.

[192] Ischebeck A, Zamarian L, Siedentopf C, Koppelstätter F, Benke T, Felber S, Delazer M. How specifically do we learn? imaging the learning of multiplication and subtraction. Neuroimage. 2006;30(4):1365–1375.

[193] Garrett D, Peterson D, Anderson C, Thaut M. Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2003 June;11(2):141–144.

Appendix A

Additional Material Included in the Comprehensive Review of the Hybrid Brain-Computer Interfacing Literature

A.1 Methods

English peer-reviewed journal articles published between January 2008 and November 2014 were chosen as the target of this review. Three major science and engineering bibliographic databases were queried: Web of Science (seven databases), Engineering Village (Compendex and Inspec) and SCOPUS. In addition, reference sections from the selected articles were scanned to find similar journal articles that might have not been found in the database queries. The last search was run on 17 November 2014.

The articles were included or excluded based on the criteria presented in Table A.1 and using the following search terms to query the databases:

1. Brain*Computer Interfac*
2. Brain*Machine Interfac*
3. Body*Machine Interfac*
4. Physiological Computing
5. Human*Computer Interaction
6. BCI*
7. Hybrid
8. Multimodal
9. EEG
10. Electroencephalogra*

Table A.1: Inclusion and exclusion criteria.

Inclusion criteria	Exclusion criteria
▪ Data collection and analysis of multimodal data including at least one neurophysiological modality, or of unimodal neurophysiological data including at least two different neural response patterns, with the aim of achieving a true hybrid brain-computer interface. ▪ Development and testing of a hybrid brain-computer interface paradigm with or without a specific application in mind.	▪ Studies using invasive modalities (ECoG, intracortical EEG, etc.) or high-cost non-portable modalities (fMRI, MEG, PET, etc.). ▪ Proofs of concept that do not include a neurophysiological modality in their analysis. ▪ Articles presenting software tools for BCIs. ▪ Review articles. ▪ Studies on experimental animals.

11. *NIRS
12. Near*infrared spectroscopy
13. *TCD
14. Transcranial Doppler Ultrasonogra*

The search terms were combined with logical operators in the following way: (1 OR 2 OR 3 OR 4 OR 5 OR 6) AND (7 OR 8) AND (9 OR 10 OR 11 OR 12 OR 13 OR 14).

Eligibility assessment was done by reading the titles of the articles first. In the case where the title did not clearly indicate if the inclusion and exclusion criteria were met, the abstract was then read. Finally, during the data collection process, the full-text articles that were found to be a poor match according to the aforementioned criteria were rejected.

A data extraction sheet was developed containing different data items relevant to the articles, based on previous reviews with similar scopes. Following a first inspection of the articles with the data extraction sheet, data items were added, removed and refined, in order to allow a useful and complete characterization of hBCIs. The extraction process was then carried out once more by going through the articles again. For each article selected, 22 data items were extracted covering five categories: the type of hBCI, the study rationale, the experimental protocol, the signal processing methodology and the evaluation of the proposed system. Table A.2 presents and defines the different items for each of these categories.

The first category (type of hBCI) includes data items that characterize the architecture of a hybrid interface. Many different hybridisation paradigms can be achieved using the currently available modalities and mental activities; a thorough description of the various configurations that have been explored so

Table A.2: Data items extracted for each selected article.

Category	Data item	Description
Type of hBCI	Hybridisation paradigms	The hybridisation paradigms encapsulate two characteristics of a BCI system: the modalities used (i.e. EEG, NIRS, EMG, etc.), and the types of activity (neural response patterns) that are used to elicit different recognizable control commands. For hybrid designs which are the aim of this study, an hBCI paradigm typically consists of a combination of one or more neural response patterns (pure hBCI), or a combination of one or more modalities (mixed or physiological hBCI).
	Modality synchronisation	See Section 2.1.8.
	Type of study	Online systems process information (and eventually give feedback) in real-time, whereas offline systems are oriented toward analysing already recorded data.
	Operation mode	See Section 2.1.8.
	Number of effective hBCI commands	The number of different commands users can trigger using the complete hBCI system.
Study rationale	Study goal	What the authors aimed to demonstrate in their article.
	Hybridisation strategy goal	The targeted outcome of using a hybridisation strategy.
	Application	If any, the further end goal of the study, or how the study could be used in a practical fashion.
Experimental protocol	Software implementation	The software or toolbox that was used to program the hBCI.
	Feedback	If any, a set of subject-perceptible changes in the hBCI, often in the form of visual or auditory cues, that informs the subject on his control of the system.
	Number of subjects	Number of subjects that were used for the analysis in the study.
	BCI experience	A subject's BCI experience refers to his previous usage of a similar BCI or familiarity with a similar recording paradigm.
Signal processing	Artefact handling	Procedures or algorithms used to handle the multiple artefacts typically arising when collecting physiological signals.
	Preprocessing	Set of manipulation steps applied on the raw data to prepare it for feature extraction.
	Features	Output of the feature extraction procedure, that typically aims at better representing the information of interest contained in the preprocessed data.
	Feature selection	Methodology used for selecting the most relevant features out of the entire set of previously extracted ones.
	Feature translation	Algorithms used by the BCI to make a decision based on the extracted features.
	Fusion level	See Section 2.1.7.
Evaluation	Performance metrics	Quantitative measures of the BCI's functioning.
	Subjective evaluation	If any, a report of the users' perceived performance in their BCI usage. Typically requires subjects to fill out a questionnaire.
	Comparison of hybrid and non-hybrid BCIs	If evaluated in the study, the advantages and disadvantages of the hBCI over a non-hybrid BCI.
	Reported challenges and limitations	Future challenges to be addressed in subsequent work, as well as reported limitations of current research.

far will thus guide further investigation. Second, the data items related to the study rationale provide information on why and in which applications hBCIs might be beneficial as opposed to non-hybrid BCIs, with the aim of enabling validation of current hybrid paradigms and of possible new applications. Third, items related to the experimental protocol of the reported studies are presented to highlight common methodological features and help in the design of similar studies. Fourth, the signal processing characteristics of the selected studies, as previously described in the second block of Figure 2.1, are reported in detail to guide the interested reader through state-of-the-art techniques for hBCIs. Finally, data items pertaining to the evaluation of hBCIs are reported to allow for an unbiased and standardised comparison of future hBCI systems. Furthermore, the comparison between hBCIs and non-hybrid BCIs,

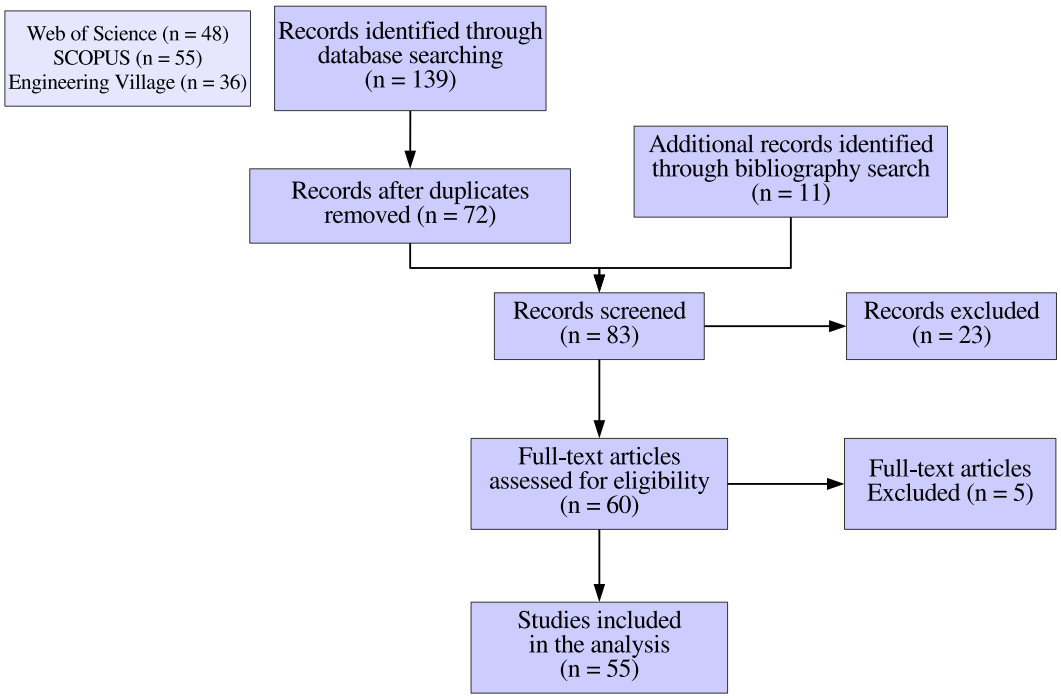

Figure A.1: Selection process for the articles from the Web of Science, SCOPUS and Engineering Village databases.

explored in some of the selected articles, is shown to validate the usefulness of hybridisation strategies. Lastly, reported challenges and limitations are presented to orient future work in the field.

One might notice the absence of a data item reporting results of each individual study, following the mentioned evaluation metrics. This information was collected, but not reported, due to the very high variability in performance metrics used from one study to another.

As summarized in Figure A.1, the database queries yielded 72 different results that matched the search terms. From these articles' reference sections, 11 additional articles were then identified. Twenty-three articles were excluded based on their title and abstract, and five more were excluded based on a full-text assessment. Therefore, 55 articles were selected for inclusion in the analysis.

Appendix B

Nomenclature of Extracted Features

The abbreviations used to describe the different EEG and NIRS features, as well as the NIRS artificial channels, are presented in Tables B.1, B.2 and B.3, respectively. Please refer to Sections 3.5.3 and 3.6.3 for a complete description of these features.

Table B.1: Nomenclature of extracted EEG features.

Abbreviation	Description
pwr_<*band power or ratio of bands*>**_**<*channel*>	Log-power in frequency bands and ratios
<*band1*>**_m-**<*band2*>**_**<*channel*>	Amplitude modulation rate-of-change
coh_<*band power*>**_**<*channel1*>-<*channel2*>	Amplitude coherence
pha_<*band power*>**_**<*channel1*>-<*channel2*>	Phase coherence
gfs_<*band*>**_**allCh	Global phase synchrony

Table B.2: Nomenclature of extracted NIRS features. A channel can be a source-detector pair (e.g. S1-D1) or an artificial channel (see Table B.3). The chromophore type can be either HbO, HbR or HbT.

Abbreviation	Description
mean_<*channel*>**_**<*chromophore*>	Average amplitude
peak_<*channel*>**_**<*chromophore*>	Latency of maximum amplitude in window (in seconds)
vall_<*channel*>**_**<*chromophore*>	Latency of minimum amplitude in window (in seconds)
diff_<*channel*>**_**<*chromophore*>	Difference between average amplitude of second and first half of window
fit1a_<*channel*>**_**<*chromophore*>	First order coefficient of linear fit of the window
fit1b_<*channel*>**_**<*chromophore*>	Zero order coefficient of linear fit of the window
fit2a_<*channel*>**_**<*chromophore*>	Second order coefficient of quadratic fit of the window
fit2b_<*channel*>**_**<*chromophore*>	First order coefficient of quadratic fit of the window
fit2c_<*channel*>**_**<*chromophore*>	Zero order coefficient of quadratic fit of the window
fit3a_<*channel*>**_**<*chromophore*>	Third order coefficient of cubic fit of the window
fit3b_<*channel*>**_**<*chromophore*>	Second order coefficient of cubic fit of the window
fit3c_<*channel*>**_**<*chromophore*>	First order coefficient of cubic fit of the window
fit3d_<*channel*>**_**<*chromophore*>	Zero order coefficient of cubic fit of the window

Table B.3: Nomenclature of artificial NIRS channels. Refer to Section 3.6.1 for the definition of each brain region.

Abbreviation	Brain region
pf	Prefrontal
lf	Latero-frontal
cf	Centro-frontal
te	Temporal
ce	Central
tp	Temporo-parietal
frM	Frontal midline