Classification multilabels à partir de signaux EEG d'imaginations motrices combinées : application au contrôle 3D d'un bras robotique

(Multilabel classification of EEG-based combined motor imageries implemented for the 3D control of a robotic arm.)

THÈSE

présentée et soutenue publiquement le 10 janvier 2017

pour l'obtention du

Doctorat de l'Université de Lorraine

(mention informatique)

par

Cecilia Lindig-León

Composition du jury

Directeur : Axel Hutt, Directeur de Recherche à l'Inria Nancy Grand-Est.

Rapporteurs : Benjamin Blankertz, Professeur des universités à TU-Berlin-Allemange.
Ricardo Chavarriaga, Chercheur Scientifique à EPFL.

Examinateurs : Laurent Bougrain, Maître de Conférences à l'Université de Lorraine.
François Charpillet, Directeur de recherche à l'Inria Nancy Grand-Est.
Fabien Lotte, Chargé de recherche à l'Inria Bordeaux.

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

Inria Nancy - Grand Est

Abstract

Brain-Computer Interfaces (BCIs) replace the natural nervous system outputs by artificial ones that do not require the use of peripheral nerves, allowing people with severe motor impairments to interact, only by using their brain activity, with different types of applications, such as spellers, neuroprostheses, wheelchairs, or among others robotics devices. A very popular technique to record signals for BCI implementation purposes consists of electroencephalography (EEG), since in contrast with other alternatives, it is noninvasive and inexpensive. In addition, due to the potentiality of Motor Imagery (MI, i.e., brain oscillations that are generated when subjects imagine themselves performing a movement without actually accomplishing it) to generate suitable patterns for scheming self-paced paradigms, such combination has become a common solution for BCI neuroprostheses design. However, even though important progress has been made in the last years, full 3D control is an unaccomplished objective.

In order to explore new solutions for overcoming the existing limitations, we present a multiclass approach that considers the detection of combined motor imageries, (i.e., two or more body parts used at the same time). The proposed paradigm includes the use of the left hand, right hand, and both feet together, from which eight commands are provided to direct a robotic arm comprising fourteen different movements that afford a full 3D control. To this end, an innovative switching-mode scheme that allows managing different actions by using the same command was designed and implemented on the OpenViBE platform. Furthermore, for feature extraction a novel signal processing scheme has been developed based on the specific location of the activity sources that are related to the considered body parts. This insight allows grouping together within a single class those conditions for which the same limb is engaged, in a manner that the original multiclass task is transformed into an equivalent problem involving a series of binary classification models. Such approach allows using the Common Spatial Pattern (CSP) algorithm; which has been shown to be powerful at discriminating sensorimotor rhythms, but has the drawback of being suitable only to differentiate between two classes. Based on this perspective we also have contributed with a new strategy that combines together the CSP algorithm and Riemannian geometry. In which the CSP projected trials are mapped into the Riemannian manifold, from where more discriminative features can be obtained as the distances separating the input data from the considered class means. These strategies were applied on three new classification approaches that have been compared to classical multiclass methods by using the EEG signals from a group of naive healthy subjects, showing that the proposed alternatives not only outperform the existing schema, but also reduce the complexity of the classification task.

Keywords: Brain-computer interfaces, EEG, motor imagery, robotics control, multilabel approach, Common Spatial Pattern, Riemannian geometry

Résumé

Les interfaces cerveau-ordinateur (ou BCI en anglais pour Brain-Computer Interfaces) mettent en place depuis le système nerveux central un circuit artificiel secondaire qui remplace l'utilisation des nerfs périphériques, permettant entre autres à des personnes ayant une déficience motrice grave d'interagir, uniquement à l'aide de leur activité cérébrale, avec différents types d'applications, tels qu'un système d'écriture, une neuro-prothèse, un fauteuil roulant motorisé ou un bras robotique. Une technique répandue au sein des BCI pour enregistrer l'activité cérébrale est l'électroencéphalographie (EEG), étant donné que contrairement à d'autres techniques d'imagerie, elle est non invasif et peu coûteuse. En outre, l'imagination motrice (MI), c'est-à-dire les oscillations des neurones du cortex moteur générées lorsque les sujets imaginent effectuer un mouvement sans réellement l'accomplir, est appropriée car détectable dans l'EEG et liée à l'activité motrice pour concevoir des interfaces comme des neuro-prothèses non assujetties à des stimuli. Cependant, même si des progrès importants ont été réalisés au cours des dernières années, un contrôle 3D complet reste un objectif à atteindre. Afin d'explorer de nouvelles solutions pour surmonter les limitations existantes, nous présentons une approche multiclasses qui considère la détection des imaginations motrices combinées. Le paradigme proposé comprend l'utilisation de la main gauche, de la main droite, et des deux pieds ensemble. Ainsi, par combinaison, huit commandes peuvent être fournies pour diriger un bras robotisé comprenant quatorze mouvements différents qui offrent un contrôle 3D complet. À cette fin, un système de commutation entre trois modes (déplacement du bras, du poignet ou des doigts) a été conçu et permet de gérer les différentes actions en utilisant une même commande. Ce système a été mis en œuvre sur la plate-forme OpenViBE. En outre, pour l'extraction de caractéristiques une nouvelle approche de traitement d'information fournie par les capteurs a été développée sur la base de l'emplacement spécifique des sources d'activité liées aux parties du corps considérées. Cette approche permet de regrouper au sein d'une seule classe les différentes actions pour lesquelles le même membre est engagé, d'une manière que la tâche multiclasses originale se transforme en un problème équivalent impliquant une série de modèles de classification binaires. Cette approche permet d'utiliser l'algorithme de Common Spatial pattern (CSP) dont la capacité à discriminer des rythmes sensorimoteurs a été largement montrée mais qui présente l'inconvénient d'être applicable uniquement pour différencier deux classes. Nous avons donc également contribué à une nouvelle stratégie qui combine un ensemble de CSP et la géométrie riemannienne. Ainsi des caractéristiques plus discriminantes peuvent être obtenues comme les distances séparant les données des centres des classes considérées. Ces stratégies ont été appliquées sur trois nouvelles approches de classification qui ont été comparées à des méthodes de discrimination multiclasses classiques en utilisant les signaux EEG d'un groupe de sujets sains naïfs, montrant ainsi que les alternatives proposées permettent non seulement d'améliorer l'existant, mais aussi de réduire la complexité de la classification.

Mots-clés: Interfaces cerveau-ordinateur, EEG, imagination motrice, contrôle robotique, classification multilabels, Common spatial pattern, géométrie Riemanienne

Acknowledgments

First, I will acknowledge the University of Lorraine, Inria and Loria, which provided me with a place to work and the opportunity of meeting people that have helped me to grow both professionally and personally.

Thanks to my advisers: Axel Hutt and Laurent Bougrain. Thank you for giving me the opportunity, and for all your help along this journey. Thank you Laurent for your time and instruction.

I want to thank in a special way my adviser in Mexico Professor Oscar Yáñez. I got first involved in BCI research because of him; and thereafter, he has been supporting me in many ways. Nothing of this would be possible if it was not for the effort and work that he selflessly offers to all his students.

To all the jury members. Monsieur François Charpillet, who has been accompanying this work since the beginning as scientific referent, always providing me with very useful and pertinent suggestions. To Fabien Lotte, for all his helpful hints. I deeply appreciate that you could make it to my PhD defense. In particular, to both reviewers. Thank you very much Ricardo Chavarriaga, I profoundly value all your remarks. Thank-you-thank-you Benjamin Blankertz, for all your time and careful report. I have learnt a lot from you. It has been a great honor.

Thanks to the BCI-lift project, not only for the financial support, but also for the opportunity of learning from all its members. Thank you Maureen for making it possible. Thank you Nathalie, Jelena and Camille.

An important thanks to Thomas Voegtlin, for all his support and company along these three splendid years. For everything I have learnt from him in all possible ways, for the lovely person he is. Right next to these thoughts, to Françoise and Jean-Marie. I was not felling far away from home thanks to you both.

To the NeuroSys team. Dear Tamara, you are a great example of calmness and compromise. Thank you very much for sharing your philosophy with me. Thank you as well for providing the best ERD/ERS% patterns ever, and for your patience in the recording sessions. Thank you dear Mariia, for the person you are: your great sense of humor, your principles and cleverness, I learnt a lot from you. Cher Sébastien, thank you very much for teaching me French, for helping me to find experimental subjects, and for all the funny moments. Thank you Francesco, it was always a pleasure to have your company. Also thanks to the former colleagues, Pedro, Meysam and Nicole. We didn't have much time to enjoy together, but it was good enough to get to know

how wonderful you are.

To the International Relation Department, thank you very much for the opportunity of joining you.

Last, but not least to my mother and all her support through the distance. To my brother and the immense love.

Contents

Chapter 1	
Brain-computer Interfaces	**1**

Chapter 2
EEG signals and motor imagery   **13**

Chapter 3
Robotic arm control   **37**

Chapter 4
Signal Processing   **47**

Appendices

Appendix A
Résumé Étendu

Appendix B
Time series of ERD/ERS% patterns

Appendix C
Offline Results

List of Figures

List of Tables

General Introduction

Brain-Computer Interfaces (BCI) systems replace the natural nervous system outputs by artificial ones that do not demand the use of peripheral nerves, allowing people with severe motor impairments to interact, only by using their brain activity, with different types of applications, such as spellers, neuroprostheses, wheelchairs, or among others robotics devices. [1], [2]. The use of BCIs is not restricted and they have also raised the interest for developing rehabilitation schemes aimed at stroke patients [3], [4], as well as interfaces for healthy people (e.g., in gaming applications).

There are several brain signals that can be used for BCI control. One of particular interest, since it has the benefit of favoring self-paced paradigms (i.e., an asynchronous operation mode), is known as Motor Imagery (MI); which can provide an intuitive and natural control to enable people with restricted mobility an ample interaction with their environment. Motor imagery involves the mental practice of a motor task without exhibiting any apparent movement [5]. The imagination of a movement includes similar brain regions and functions as those engaged in programming and executing a real one. The main difference is that in the former case, movements are blocked at a cortico-spinal level [6].

The EEG activity elicited during motor imageries consists of changes of sensorimotor rhythms that exhibit an oscillatory behavior that is of particular interest within two specific frequency ranges: the alpha (8-13 Hz) and beta (13-25 Hz) bands [7]–[9], since they have been shown to be modulated during and following the preparation and execution of imagined movements [10]. This modulation takes two different forms: an oscillatory power reduction during the motor task known as Event-Related Desynchronization (ERD); and an oscillatory power increase known as Event-Related Synchronization (ERS) that appears just after the end of the motor task over a few hundred milliseconds [11]. During the preparation of a motor task certain cortical and subcortical regions are activated; contrarily, in the course of motor imageries most of the activity is found within the primary motor cortex over the corresponding contralateral hemisphere [12]. The topographic representation and band power change during motor imageries are known to be located in specific regions on the primary motor cortex (broadly, left hand activities are observed over electrode C4, right hand activities on C3, feet activities on C_z, etc.) [11], [13].

The number of commands afforded by motor imagery-based BCIs depends on the amount of brain states that the system is able to detect which, in turn, is limited to the number of body parts that users are able of producing suitable patterns with. Moreover, the difficulty of such a task increases as the total of available commands grows, which is worthwhile only if a

suitable amount of effectively separable human brain states can be assigned [14]. There are two approaches for solving multiclass problems: On the one hand some techniques directly forecast the predicted label by using a straightaway method, such as distance-based algorithms; and on the other hand some alternatives apply multiclass extensions to decompose the problem into an ensemble of binary tasks that can be accomplished in different ways, for instance, by using the classical one-versus-one or one-versus-all approaches [15].

An alternative multiclass scheme that allows affording users with a plenty repertory of different commands is known as multilabel approach, which addresses the detection of combined motor imageries (i.e., two or more body parts used at the same time) [16], [17]. In comparison with the standard scheme, this approach has the advantage of considerably increasing the number of different brain states while using the same number of body parts (in order of 2^P compared to P, where P is the number of body parts).

In contrast to the standard multiclass paradigms, upon which most of the research has been devoted, less work has been reported about brain oscillatory patterns induced by combined motor imageries. The empirical mode decomposition technique in combination with phase synchronization analysis, has been proposed and applied to evaluate the human brain synchrony of the supplementary motor and the primary motor areas during complex motor imagination of combined body and limb action, from where the classification performance suggested the feasibility for BCI applications in motor nerve reconstruction of lower limbs [18]. In addition, aiming at overcoming the limited amount of output commands resulting from the scarce number of classes, and for the purpose of achieving continuous 3D control, both hands motor imagery was adopted to complement the set of instructions in a simple limb motor imagery based-BCI [17]. More recently, the characteristics of EEG patterns during simple and combined motor imageries were investigated to confirm the separability of seven different classes for motor imagery based-BCI systems implementation [16]. Another study reported that the movement of several fingers at the same time is not the simple superposition of individual finger movements at the ECoG scale, but rather the activation of additional neural networks. Which implies that for BCI implementation purposes, the generation of control signals encoding complex movements would likely use signal components beyond individual members superposition [19]. However, to our best knowledge the latter conclusion remains to be confirmed at the EEG scale; which has become the main interest of the present work.

The importance of this question lies on the particular interest of optimizing the Common Spatial Pattern (CSP) algorithm for its application on multilabel classification tasks. This method is a very convenient and widely used algorithm to extract motor imagery features from EEG signals. It generates a series of spatial filters that are applied to decompose multi-dimensional data into a set of uncorrelated components [20], which aim at extracting elements that simultaneously maximize the variance of one class, while minimizing the variance of the other one. Because of the manner in which the CSP algorithm achieves such a contrast, this approach has the drawback of being suitable only to discriminate between two classes and, in order to solve problems involving

many conditions, multiclass extensions are required [2], [15]. Classical extensions such as the one-versus-one and one-versus-all approaches present several disadvantages. Perhaps the most critical one is the extensive amount of models that is required as the number of classes increases, which negatively impacts on the computation time and on the reliability of results. In this sense, by confirming that the motor imagery of a combination of two or more body parts involved at the same time corresponds, in fact, to the simple superposition of the individual members at the EEG scale, it is possible to design more convenient approaches for adapting not only the CSP method but any other, since, under this scheme, it becomes possible to use an ensemble of binary classifiers to determine at each activity source whether or not its associated body part is engaged in the task, and thus to infer the final decision from the combination of all subproblems [21], [22].

In the present study we provide some evidence suggesting that this hypothesis is plausible, upon which we have developed three new feature extraction methods for discriminating combined motor imageries, namely, One-step Multilabel (OsM) approach, Hierarchical Multilabel (HM) approach, and One-step Hierarchical Multilabel (OsHM) approach, which to our best knowledge have not been considered before. The motivation was to provide a full 3-D control for a commercial robotic arm that affords users with fourteen predefined motions. The proposed paradigm involves the use of three body parts, namely, right hand, left hand and feet; which in combination with the rest condition, constitutes an 8-class problem. For the implementation, a switching-mode control that allows managing different actions by using the same command was proposed. Six healthy subjects participated in one recording session comprising three runs of ten trials per class, which totals 30 trials per class. Results were slightly superior as those obtained by the classical one-versus-one and one-versus-all approaches, which is of particular interest if we consider that, in order to solve the aforementioned 8-class problem both methods demand, respectively, 28 and 8 feature extraction models; which is by far more than the 3, 7 and 6 models that are required by the OsM, HM and OsHM approaches.

The remainder of this document is organized as follows: The first chapter is dedicated to provide a general basis of the principles and practices of Brain-Computer Interface (BCI) systems. These aspects will be further deepened along the second chapter in the context of motor imagery. Along these sections we will present the 4-*class* database, which comprises a series of EEG recordings from five healthy subjects who participated in a motor imagery experiment involving the use of the right hand, left hand and both hands together; which, in addition with the rest condition, leads to a paradigm of four different classes. The 4-*class* database has been created to carry out a preliminary study to assess the similarities of the activity that is found over the electrodes associated with the use of the left and right hands during simple and combined motor imageries. In this way, by proposing a small number of classes, it is possible to register a considerable amount of trials without causing users fatigue. Upon these data we will present a series of statistical results based on the analysis of the power spectrum observed at each electrode during the four different motor tasks, which supports the hypothesis that EEG data recorded during a combined motor imagery can be characterized as the superposition of the activity that

is independently generated by each one of the sources associated to the body parts engaged in the task. Subsequently, along the third chapter, we will describe the characteristics of the robotic arm that has motivated the practical solutions developed throughout this project. In the same line, we will describe the operation scheme that has been designed; which consists of a smart implementation based on a switching-mode control, that allows managing different actions by using the same command. As mentioned, this device affords users with fourteen predefined motions, which demands far more commands than those obtained with the 4-*class* database. With this in mind, we have generated the 8-*class* database, which comprises EEG signals from 6 healthy subjects inducing the eight mental states that it is possible to generate by the motor tasks involving the use of the right hand, left hand, both feet together, and rest. Similarly as with the 4-*class* database, we will present a qualitative analysis of the ERD patterns that are found in the EEG signals during the eight different tasks, together with a statistical comparison of the power spectrum at each electrode among each pair of conditions. These results create the basis upon which we will justify the grouping of classes that has been suggested for the feature extraction methods contributed by the present work. In the fourth chapter the concept of feature extraction is introduced and elaborated in the context of EEG signals during motor imagery. Throughout these pages, the CSP algorithm will be elucidated together with two of its variants, namely, Analytic Common Spatial patterns (ACSP) and Common Spatial Pattern by Joint Approximate Diagonalization (CSP by JAD). Additionally, we will present a strategy called Filter Bank Common Spatial Pattern (FBCSP) that optimizes the algorithm performance by decomposing the EEG signals into a series of frequency bands, which favors the selection of a discriminative subject-specific frequency range. Subsequently, we will present the notion of classification, for which two commonly used algorithms will be described, namely, Linear Discriminant Analysis (LDA) and Support Vector Machines (SVM). Moreover, in order to explore different solutions, the present work considers as well the use of Riemannian geometry; with which classification can be directly achieved within the manifold of symmetric and positive definite matrices that is given by the covariance matrices of the EEG signals. Afterwards, in the last section of this chapter, we will introduce a new classification method for motor imagery-based BCI systems that makes use of the combination of the CSP algorithm and Riemannian geometry, which represents one of the contributions of the present work. In the fifth chapter the concept of multiclass approach will be introduced by describing three classical methods, namely, one-versus-one, one-versus-all, and hierarchical algorithms. At this point we will describe the three new multilabel methods that we have implemented, namely, One-step Multilabel (OsM) approach, Hierarchical Multilabel (HM) approach, and One-step Hierarchical Multilabel (OsHM) approach. The basis of each one of these new methods lies in the results that were presented along chapters 2 and 3, upon which we have proposed a convenient manner to carry out data partition, so that the entire problem is reduced into a series of binary classification tasks. Finally, along the sixth chapter we will present the results, in terms of accuracy, that were obtained by each one of the described methods over both databases. The corresponding discussion is presented in chapter 7 and, the general conclusion, in

chapter 8.

1

Brain-computer Interfaces

Contents

Brain-computer interface (BCI) systems consist of a combination of modules that allow recording the user brain activity in order to analyze it and find specific patterns indicating that the person is executing a particular mental task, which in turn is associated to a specific command for controlling the function of certain devices, such as computers, robotics, wheelchairs, prostheses,

orthoses, etc. In this way, BCIs replace the natural nervous system pathways by artificial ones, which allow people with a severe motor impairment to interact with their environment by using only their brain activity. Most of BCI applications aim at improving the quality of life of their users, who have lost certain function(s) due to a degenerative disease or an accident, in which case the system will attempt to restore the missing ability. There are other functions that BCIs undertake according to the user needs and system scopes, which will be presented in the following chapter.

The initial part of this document is dedicated to introduce the notion of BCI systems. To this end, their general architecture will be presented along the first section, where the function of each of their components will be described. Additionally, the concepts underlying the BCI terminology, as well as the different types of brain activity that are suitable for their design, will be detailed. Finally, we will present the basic applications that BCIs are aimed for based on the nature of the brain activity that is used to design different paradigms. In this sense, we will emphasize on Motor Imagery (MI) since, as it will be further exposed, its nature facilitates robotics control.

1.1   BCI definition

Brain–Computer Interfaces consist of a direct communication pathway between the brain and an external device. BCIs are often directed at researching, mapping, assisting, augmenting, or repairing human cognitive or sensory-motor functions. A recent definition involving all usages of BCI systems is stated as follows (see Figure 1.1):

A BCI is a system that measures central nervous system activity and converts it into an artificial output that replaces, restores, enhances, supplements, or improves natural outputs and thereby changes the ongoing interactions between the central nervous system and its external or internal environment [1].

1.2   BCI architecture

BCI systems can be presented in different variations, however, from a general point of view there are three main components that every BCI includes in its architecture, namely, a signal acquisition module, a signal processing stage, and the corresponding output, which can presented in many different forms according to the application of the system [1], [23]–[25]. In addition, BCIs might include a feedback module to provide users with some specific information, such as brain signals and/or classifiers outputs. Figure 1.1 shows the general framework of a BCI where it is possible to observe each one of the aforementioned modules.

[Wolpaw & Wolpaw, *BCI: Principles and Practice*].
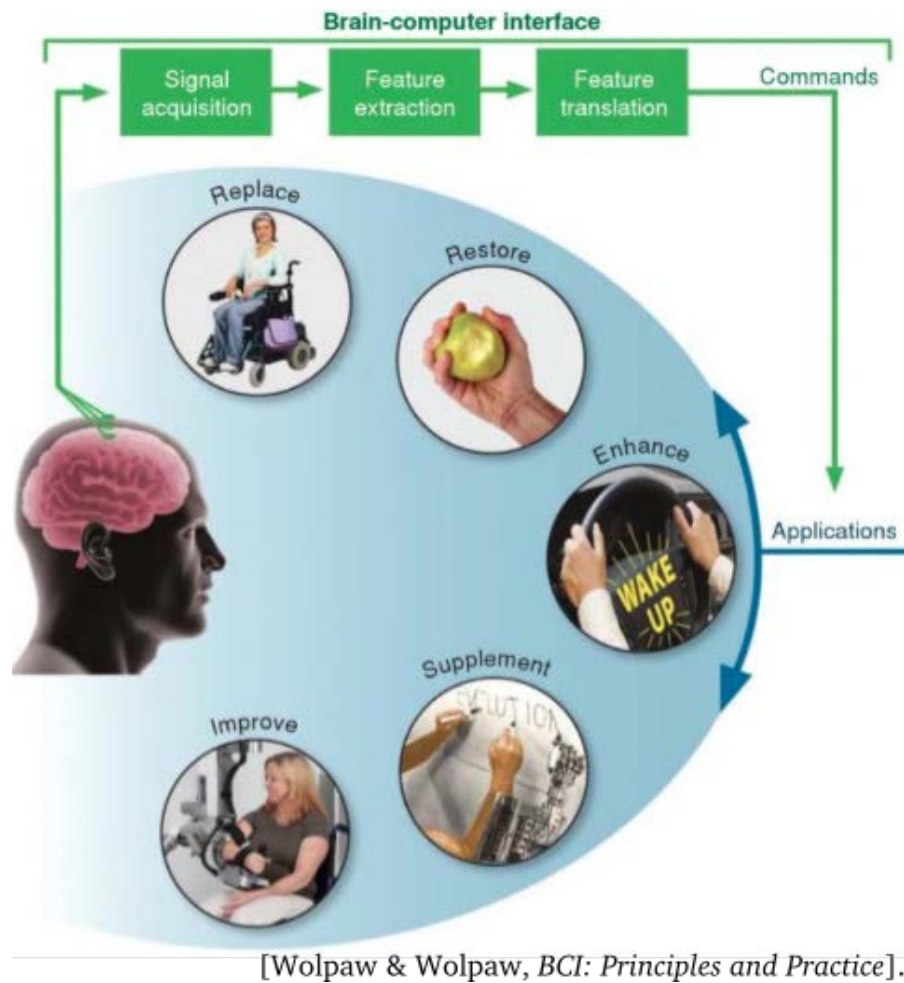
Fig. 1.1: General BCI architecture. There are three main components that every BCI includes in its architecture, namely, a signal acquisition module, a signal processing stage, and the corresponding output, which can presented in many different forms according to the application of the system.

1.2.1 Signal acquisition

As it will further seen along this chapter, brain signals can be recorded at different scales depending on whether the system is invasive (i.e., direct contact with the cortex) or not (superficial contact onto the skin). In both cases electrodes are located over the regions of interest and the acquired signals are amplified and digitized to be used by the BCI system. In this stage some preprocessing steps can be considered, such as the use of a notch filter to suppress the line noise, and/or the use of band-pass filters to extract the frequencies of interest.

1.2.2 Signal processing

Signal processing is the analysis of the brain signals that is carried out in order to extract and interpret useful information related to the user intention. Basically, there are two main aspects

of signal processing in the context of BCI, namely, feature extraction and feature translation. The first concept consists in determining the characteristics that confer distinctiveness to the brain activity that is elicited during the corresponding task, so that it can be recognized from other states. On the other hand feature translation consists in the classification of the extracted features in conjunction with the assignment of the corresponding action.

1.2.2.1 Feature extraction

The goal of feature extraction is to select relevant information from the recorded signals in order to fit a classification algorithm (classifier) that is able to categorize such elements according to their position within the classification space, so that the system can proceed to translate the corresponding command.

In most of the cases this process reduces the dimensionality of data by excluding redundant information, and it enhances the attributes of the signal by exploiting the aspects that are related to the physiological phenomena. The derived values are projected into the feature space, where elements from different conditions are expected to be located as separated as possible, so that distinct groups can be easily identified.

1.2.2.2 Feature classification

The aim of classification is to assign the feature vectors to one of K discrete classes, where K corresponds with the number of different brain states that the system is able to detect and to translate into commands for controlling the interface.

The classification algorithm can be implemented for distinguishing between two classes (binary problems), or several classes (multiclass problems). In this sense, given that the difficulty of the classification task increases with the number of classes, there is a trade-off between providing a natural and intuitive control by affording users with many commands, or ensuring an effective performance with a limited action.

1.2.3 System output

As mentioned, the number of considered mental states (classes) are associated to control actions (commands). These actions can be used to control an unlimited number of different types of applications, the most common ones are based on the use of devices such as computers for communication or control, electric wheelchairs, and robotic limbs to be used in rehabilitation or for accomplishing motion actions. Some applications include movement restoration, in which BCI systems bypass a spinal cord injury and activate muscles with electrical stimulation.

1.2.4 Feedback

In many cases, BCI systems can provide useful information that may be used to improve their control. For instance, o detect and inform users about execution errors. This feature is not always

included in BCI applications, even though it can be specially useful during on-line sessions, since errors may be critical and it becomes important to inform users about their progress. Recently, BCI researchers attempt to incorporate feedback as a tool for training and developing users skills [26].

1.3 BCI taxonomy

Aside from the general architecture shared by BCI systems, there are other aspects from which specific features can be distinguished. Such as the nature and manner in which the targeted brain activity is induced, the operating protocols and the type of decoding which, among other characteristics, provide a considerable number of different alternatives for BCI applications that can be pursued according to the users' needs. In the following sections we provide a description of some important features for BCI design.

1.3.1 Dependent and independent BCIs

The terms *dependent* BCI and *independent* BCI were recently introduced [27], according to the basic definition, both use brain signals, but the differ in their dependence on natural outputs induced by the central nervous system. In this sense, a dependent BCI uses brain signals that depend on muscle activity. For instance, the P300 speller demands user to direct their gaze towards the screen position where the symbol they want to select is located. Even though dependent BCIs do not provide the brain with a new output that is independent of natural outputs, they may still be useful [1].

On the contrary, an independent BCI does not depend on natural outputs at all, since muscle activity is not fundamental for inducing the brain signals that are used by the BCI. For instance, in motor imagery-based BCIs, users can modulate their brain signals by imagining a motor task in order to control the system output. For people with severe motor impairments independent BCIs are likely to be more useful. However, it is important to acknowledge that nowadays most of the BCIs are neither completely dependent nor completely independent. Just as it is necessary to control gaze direction to use the P300 speller, most of motor imagery-based BCIs require the user having sufficient visual function (and thus gaze control) to watch the results of the BCIs output commands (e.g., cursor movements).

1.3.2 Exogenous and endogenous BCIs

Brain activity for BCI control can be elicited spontaneously as the result of an internal process, or can be triggered by an external stimulus. The former case receives the name of endogenous brain activity, which can be induced by some mental tasks such as the imagination of body movements. The second case is known as exogenous activity and can be evoked by either visual, auditory or somatosensory stimuli intended to induce patterns that are easily detectable.

Endogenous BCIs are likely to provide more freedom. For example, an endogenous BCI can

allow users to move a cursor towards any direction in a 2-dimensional (or even 3-dimensional) space, whereas an exogenous BCI can allow subjects to make a choice between different symbols located on the same space. However, an important advantage of exogenous BCIs is that they do not require users to be trained, and thus they can be effectively used since the very first time, contrarily to endogenous BCIs, which sometimes require long periods of training, without implying that they would work accurately afterwards.

1.3.3   Passive and active BCIs

Another recent term, *passive* BCI, is applied to BCI applications that use brain signals associated with aspects of the user's current state, such as level of attention [28]. For instance, a BCI might detect EEG features preceding lapses in attention and produce an output (e.g., a sound) that alerts the user and restores attention. In this sense, the term *passive* is intended to distinguish between this kind of BCI applications from those that provide communication and control (i.e., *active* BCIs). However,*passive* and *active* are subjective terms that lack clear neuroscientific definitions. Moreover, continued use of a passive BCI might induce adaptations of the central nervous system that improve its performance, so that the term *passive* becomes no longer applicable. Thus, it seems preferable to categorize BCI applications simply as shown in Figure 1.1, in which case passive BCIs will generally fit into the *enhance* or *supplement* category.

1.3.4   Hybrid BCIs

The concept of *hybrid* BCI can be applied in two different ways [29]. It may point to a BCI that uses two different kinds of of brain signals (e.g., visual evoked potentials and sensorimotor rhythms) to generates its outputs or, alternatively, it may describe a system that combines a BCI output with a natural muscle-based output. In the latter usage, the BCI output supplements a natural output (e.g., as illustrated in Figure 1.1).

1.3.5   BCI operating protocols

As with any other task, a BCI user may by sometimes actively using the BCI (e.g., using it to communicate with someone), or may be not using it at other times (e.g., is doing something else, sleeping, daydreaming, etc.) [23], [24]:

— The state of active BCI use is referred to as *intentional control*.

— The state of other activities is referred to as *no intentional control*, or more simply, *no-control*.

Ideally, BCI use should be as natural and convenient as a person's muscle-based actions, so that the BCI is able to operate in a self-paced mode, in which the system is readily available and

users can achieve *intentional control* when desired. To afford such user-paced operation, a BCI must be able to reliably distinguish between the user's *no-control* state and the user's *intentional control* state. This is a difficult requirement because a wide range of brain activity signals might correspond to *no-control*, and the system must be able to recognize all of them as indicating such state. If the BCI is not able of accurately distinguishing between *no-control* and *intentional control*, brain activity during *no control* will sometimes be translated into unintended actions. For many BCI applications, even a small number of such false-positive errors could make BCI use very frustrating and possibly impractical.

Currently, BCI operating protocols address the need of avoiding false-positive errors in one of two ways: Trough *synchronous* protocols, which limit BCI actions to times when the user can safely be assumed to be in the intentional-control state; and, alternatively, through *self-paced* (or *asynchronous*) protocols, which are able to distinguish satisfactorily between the *no-control* state and the *intentional-control* state.

1.3.5.1 Synchronous protocols

As mentioned, *synchronous* protocols limit the times during which the BCI will actually translate the user's brain activity into an action [23], [30]. A common approach for providing synchronous protocols consists in giving the user a specific sensory stimulus and to analyze the brain activity that is induced during a fixed period immediately afterward. In this way, the system converts brain signals into outputs only during that poststimulus period, so that it limits and specifies the times when the BCI will convert brain activity into an action. Alternatively, the protocol might provide cues to inform users of an impending period of expected *intentional control*.

1.3.5.2 Self-paced protocols

A *self-paced* protocol (also referred to as an *asynchronous* protocol) is able to distinguish between no-control and intentional-control states. This mode of operation allows users to control the BCI whenever they desire. Self-paced protocols can afford more natural and dynamic interactions between the user and the BCI, but they are not yet suitable for general out-of-the-lab use.

Self-paced BCIs can rely on motor imagery as well. For instance, if the system detects the EEG activity associated with one of several specific mental tasks (such as imagining a left-hand movement), then that activity is translated into a particular command.

1.3.6 Continuous and discrete decoding

In general, BCI systems have been used to provide one of two types of decoding: *continuous* or *discrete*. Continuous decoding is best described as real-time control of an ongoing movement, such as of a cursor or a robotic limb. Contrarily, discrete decoding consists of classifying specific events, as when a letter or icon is selected from among a set of possible choices.

1.4 Brain Signals for BCI

Most of the activity for BCIs implementation purposes consists of brain's electrical signals, which can be recorded at different scale levels, from the centimeter scale of electroencephalography (EEG) through the millimeter scale of electrocorticography (ECoG) to the tens-of-microns scale of neuronal action potentials (spikes). Which recording resolution is more advantageous for which purposes remains as a research topic, since even though the activity recorded from basic units suggests that spikes could provide several degrees of freedom and might thus be the best signals for BCIs use, the evidence that adaptive methods can elicit multiple brain states also from EEG signals, points out that the difference between BCI performance provided by these different scales may not be as nearly important as the difference in their topographical resolutions.

Recently, nonelectrical metabolic signals, such as those generated by functional near-infrared spectroscopy (fNIRS) and functional magnetic resonance imaging (fMRI), have drawn the attention of BCIs developers due to their potentiality of exposing suitable brain activity patterns. On the one hand, fNIRS has the advantages of being noninvasive and inexpensive, whereas on the other hand, despite being much more expensive, fMRI is not invasive either and offers a much higher resolution.

Besides of the suitability of brain signals for BCI purposes, a very important aspect consists of the access to the corresponding sources, which may demand an invasive procedure that constrains the feasibility of certain techniques. In the following sections, some examples of invasive and non-invasive recording methods are provided together with their advantages and disadvantages.

1.4.1 Invasive recording techniques

Invasive recordings are often used in humans only for evaluation purposes in patients suffering from epilepsy, Parkinson's disease or others neurological conditions. In these cases electrodes are implanted temporarily (5-14 days [31]) to evaluate the possibility of undergoing surgery for the extreme cases where patients are resistant to pharmacological treatments [32]]. The main goal of this technique is to identify the brain tissue where seizures are originated. To this end, electrodes are directly implanted on or in the cortex through a surgical procedure, which provides a high temporal and spatial resolution. In general, BCI research by using invasive recordings methods are based on animals, since humans are mostly available only as patients undergoing a surgical procedure who accept to collaborate in the research topic and, in many cases, the implanted electrodes are not located over regions that are of interest for BCI purposes.

Invasive techniques can be subdivided into two major kinds of recordings regarding the type of electrodes that are implanted, namely, electrocorticogram and intracortical recordings.

1.4.1.1 Electrocorticography

Electrocorticography (ECoG) recordings, also called intracranial EEG (iEEG), use an array of electrodes (on a subdural grid) that is placed in direct contact with the surface of the cortex

without penetrating the brain. ECoG electrodes are usually arranged in a matrix whose size depends on the region that needs to be covered. This technique offers a high spatial resolution and a high signal-to-noise ratio due to the proximity with which neuronal activity is registered. However, it also presents important inconveniences, such as those related to the required surgery and the following recovery, which also difficult the possibility of a subsequent rearrangement of the electrodes position.

1.4.1.2 Intracortical Recordings

Microelectrodes can be placed close (i.e., extracellular recordings) or within the neurons (i.e., intracellular recordings). Usually BCI studies based on both types of techniques are restricted to animals. Signals recorded at this scale are action potentials, also referred to as local field potentials in the case of intracellular recordings, or to as spikes for extracellular recordings.

— **Neuronal Action Potentials (Spikes)**: Consist of short electrical pulses emitted by single neurons during information transmission down to the axon. Such recordings can be achieved by placing electrodes close to the cell or to the axon. When only the activity of one neuron is observed, registers are referred to as single-unit recordings. In contrast, when electrodes are in contact with more than one neuron, recordings are referred to as multi-unit. A "spike train" or "temporal coding" corresponds to a series of spikes generated by a neuron either spontaneously, or in response to an external stimulus.

— **Local Field Potentials (LFP)**: Reflect the activity that is generated by a population of neurons, which can be measured through extracellular electrodes. Neurons in the neighborhood generate overlapping spikes, which causes a large amplitude of LFPs, evidencing the synchronous activity between neurons. This is useful to understand their behavior to a specific stimulus.

1.4.2 Non-invasive recording techniques

Non-invasive techniques are distinguished for recording brain activity in a superficial manner (i.e., without penetrating the skin), so that the process is painless and safe for users. Compared to invasive recording methods, non-invasive alternatives offer several advantages; since they do no represent any risk for their users, such as the hemorrhages, infections, or/and bioincompatibility that may be caused by implanted electrodes, without forgetting the hospitalization, recovery, and the trauma induced by the required surgery. Because of this reason, non-invasive techniques are by far more popular for BCI purposes. Some of these methods include functional Magnetic Resonance Imaging (fMRI), functional Near Infrared (fNIR), magnetoencephalography (MEG), and electroencephalography (EEG).

1.4.2.1   Functional Magnetic Resonance Imaging (fMRI)

Blood flow and oxygenation changes are closely related to neural activity [33]. Based on these principles, fMRI is intended to measure changes in the blood flow for detecting active brain regions. Basically, when neurons go active, blood flow carries glucose and oxygen. Oxygen is transported by hemoglobin in the form of oxyhemoglobin, which presents magnetic properties. These cells quickly degrade the local oxygen and increase the level of deoxyhemoglobin (hemoglobin without oxygen). The advantage of this technique is its high spatial resolution (1 mm^2), though it presents a low temporal resolution (5-8 seconds), given that inflow of blood is not instantaneous. In addition, this technique is cumbersome and expensive.

1.4.2.2   Functional Near Infrared

fNIR measures the absorption of near infrared light (NIR) in the brain for monitoring changes in tissue oxygenation. It offers a suitable spatial resolution (1 cm^2), but a relatively low temporal resolution, given that determination of changes in the oxygenation of the tissues can take from 5 up to 8 seconds. fNIR is a portable system, since sensors are placed in the prefrontal cortex by using a sort of "headband", which provides users with comfort. It consists of an emerging technology that is similar to fMRI, but it overcomes many of its drawbacks by being affordable, portable and wearable. However, despite these advantages, fMRI has better spatial resolution, and it offers the possibility of obtaining an image of the entire brain [34].

1.4.2.3   Magnetoencephalography (MEG)

It is based on measurements of the magnetic fields generated by the electrical activity produced in the brain [35]. This technique favors a high spatial (millimeters) and temporal (milliseconds) resolution, in addition to a relatively high signal-to-noise ratio (SNR). However, the equipment is expensive, cumbersome and imposes special requirements, such as a magnetically shielded room to reduce artifacts generated by electrical currents or ferromagnetic objects. Despite its drawbacks MEG can be a useful tool for brain-computer interfaces.

1.4.2.4   Electroencephalography (EEG)

Given that this technique is safe, easy to use, affordable, and offers a high temporal resolution (in the order of milliseconds), it has become the most common method for recording brain signals in humans. Roughly speaking, EEG records the electrical activity generated within different regions of the brain. EEG recordings use a set of electrodes that are positioned on specific locations over the scalp. The majority of electrodes require a conductive gel for improving signals transmission, though recently some varieties offer the alternative of "dry electrodes", which facilitates their use and result more comfortable for users. EEG electrodes can be either *passive* or *active*. The former case refers to sensors that are connected to an amplifier by using a cable, whereas the latter type include the amplifier in their taxonomy, which prevents from amplifying the additional noise

that is present in the environment and/or caused by movements in the cable. Additionally, some brands offer wireless EEG systems that allow complete mobility to their users. Even though the spatial resolution of this kind of systems is not as convenient (at best 1-2 cm), and considering that they present a low signal-to-noise ratio, this solution still offers several advantages for a daily-based use.

2

EEG signals and motor imagery

Contents

Along this chapter we will provide a framework of the principles and use of electroencephalographic (EEG) recordings in the context of motor imagery-based BCIs. The first section aims at describing the international 10-20 system of electrode placement since, during the effectuated experiments, sensors location was based on such standard. Subsequently, the different frequency bands that are of interest for the analysis of EEG signals will be introduced, together with the regions over

the motor cortex that are activated during a motor task. In this regard, the physiological aspects that take place within the aforementioned areas, such as Event-Related Desynchronization (ERD) and Event-Related Synchronization (ERS), as well as their lateralization, will be elucidated. Furthermore, we will develop the concept of oscillatory power by providing its definition and a computation example, which will be useful to visualize the ERD/ERS patterns that are expected during a motor imagery task.

As a consequence of this review, we will point to the fact that the amount of provided commands to control a motor imagery-based BCI, depends on the different brain states that the system is able to detect; which, in turn, is limited to the number of body parts that users are able of producing suitable patterns with. In this sense, we will introduce the concept of *multilabel* approach; which dramatically increases the quantity of afforded commands by making use of the combination of two or more body parts at the same time, and thus favoring an intuitive and natural control of the system. At this point we will introduce the main hypothesis of the present study; which states that EEG data recorded during a combined motor imagery can be characterized as the superposition of the activity that is independently generated by the sources associated to the body parts engaged in the task. In order to verify this premise, we will introduce a database, to which we will refer as "4-class database", comprising the EEG recordings of five healthy subjects performing three different motor imageries, namely, right hand, left hand and both hands; which, together with the rest condition, sets up four different brain states. Upon these data we will present a series of statistical results based on the analysis of the power spectrum observed at each electrode during the four different motor tasks that supports the aforementioned hypothesis, as well as a qualitative study of the ERD patterns that are generated under the considered conditions.

From the confirmation of this hypothesis, we will develop in the following chapters a new strategy for multilabel classification; which consists in one of the contributions of the present work.

2.1   EEG signals

Most of BCI applications use electric or magnetic signals produced by the brain activity in order to transfer the user's intention to the system, as in this way it is possible to obtain recordings noninvasively from sensors on or above the scalp. The two main noninvasive extracranial techniques used for BCI purposes are electroencephalography (EEG) and magnetoencephalography (MEG), from which EEG is the most widely used considering that it is inexpensive, convenient, and adaptable to diverse environments; thereby, has been developed commercially into portable and even wireless models for a broad diversity of purposes.

The underlying phenomena that is observed through EEG signals is caused by the activity of neurons, which are inherent polarized by membrane transport proteins that pump ions out and into the cells. The exchange of ions occurs constantly between neurons and the extracellular

milieu, either to maintain the resting potential or to propagate action potentials. Ions with the same charge repel to each other, and when many of them are pushed out at the same time from different neurons, they can push their neighbors, who push their neighbors, and so on, which generates an electric wave. When such a wave reaches the electrodes on the scalp, its ions can push or pull electrons on the metal surface of the sensors. Since metal easily conducts the exchange of electrons, the difference of voltage between any two electrodes can be measured over time, which can be registered as EEG recordings. The electric potential produced by an individual neuron is far too small to be registered; therefore EEG activity always reflects the synchronous activity of thousands or millions of a network of neurons having similar spatial orientation. When neurons are not similarly oriented, their ions do not contribute in creating a measurable wave.

Most of the EEG activity is believed to be produced by pyramidal neurons within the cortex; since they are well aligned and fire together. However, voltage magnitude falls as the distance between the source and sensors increases due to the volume conduction phenomena, which points out to the loss and spread of currents across the different physiological tissues during their transit to reach the electrodes, so that activity from deep sources is more difficult to detect that the currents generated near the skull. EEG activity shows oscillations at a variety of frequencies, which present characteristic ranges and spatial distributions that are associated with different brain states.

2.1.1   EEG brain rhythms

EEG patterns are typically characterized in terms of rhythmic activity and transients. Rhythmic activity consists in the synchronized oscillation of a population of neurons that fire together at certain frequency within a particular region, which allows making the link between certain physiological processes and their associated brain function, opening the possibility of using such modulations as interpretable commands to control BCI applications. Most of the EEG activity under standard clinical protocols is found within the range from 1 to 30 Hz, components that are observed below or above this range are likely to be artifacts.

In the context of EEG, rhythmic activity is separated into frequency bands wherein the observed oscillations have demonstrated to be associated with certain biological significance. The range and nomenclature of these bandwidths is stated as follows (see Figure 2.1):

— **Delta rhythm** (< 4 Hz): It is a normal high amplitude wave detectable in the EEG recordings of healthy subjects. In adults it appears within frontal areas whereas, in the case of children, it is present over posterior regions. Delta waves are commonly found during deep sleep, dreamless sleep, trance, deep hypnosis and, rarely, during wakefulness. In this sense their presence has been reported along some continuous-attention tasks [36]. This rhythm is affected by drugs, alcohol, anxiety, depression and other factors that directly or indirectly disturb sleep. Pathological signs might be related with sub-cortical, midline or diffuse lesions, as well as with metabolic encephalopathy hydrocephalus.

— **Theta rhythm** (4 - 7.5 Hz): It is found within the parietal-temporal area during drowsiness, REM (i.e., Rapid Eye Movement) sleep, and in the transition from sleep to wakefulness, with higher amplitude in young children. It is also elicited by emotional stress, alcohol ingestion and fatigue [37]. Furthermore, it has been associated with inhibition of elicited responses, since it increases in situations when a person is actively trying to repress certain reactions [36]. In pathological cases it is associated with focal subcortical lesions, metabolic encephalopathy, deep midline disorders, and some instances of hydrocephalus. The theta rhythm can be characterized into two types: the hippocampal theta rhythm, observed in the hippocampus of mammals (e.g., cats), and the cortical theta rhythm usually recorded on the human scalp.

— **Alpha rhythm** (8 - 13 Hz): First defined by [38] as a rhythm found over posterior regions, generally with higher voltage over the occipital areas, during wakefulness. Power within this band is increased with closed eyes in a state of physical relaxation, over the occipital, parietal and posterior temporal regions. Alpha waves are also associated with inhibition control. Given the link between alpha waves and relaxed mental states, their power increase has become a desirable outcome for some types of biofeedback training. EEG can be used to provide the subject with feedback when alpha waves increase, enabling some individuals to consciously increase this activity.

— **Beta rhythm** (14 - 30 Hz): Beta waves are detectable in almost every healthy adult over the parietal and frontal lobes. Their presence at multiple and varying frequencies with a low amplitude is often associated with active concentration [39]. Over the motor cortex, beta activity is suppressed prior to and during muscle contractions of isotonic movements and it is frequently mixed with a mu rhythm (see below) [40]. Beta activity is increased when movements have to be resisted or voluntarily suppressed [41]. In some cases beta activity is also considered within separately ranges as lower beta band (14–20 Hz), and upper beta band (15 - 30 Hz).

— **Gamma rhythm** (> 30 Hz): This rhythm is typically associated with perception and consciousness, thought it has not fully demonstrated and, according to some scientists, their presence may be caused by electromyographic artifacts [42].

— **Mu rhythm** (8 - 12 Hz): These waves are present within the same range as the alpha activity but they differ in their topographical location and they are triggered by endogenous responses. Mu activity decreases during movements or preparation of movements and increases with relaxation. These phenomena are known, respectively, as *Event-Related Desynchronization* (ERD) [43] and *Event-Related Synchronization* (ERS) [3].
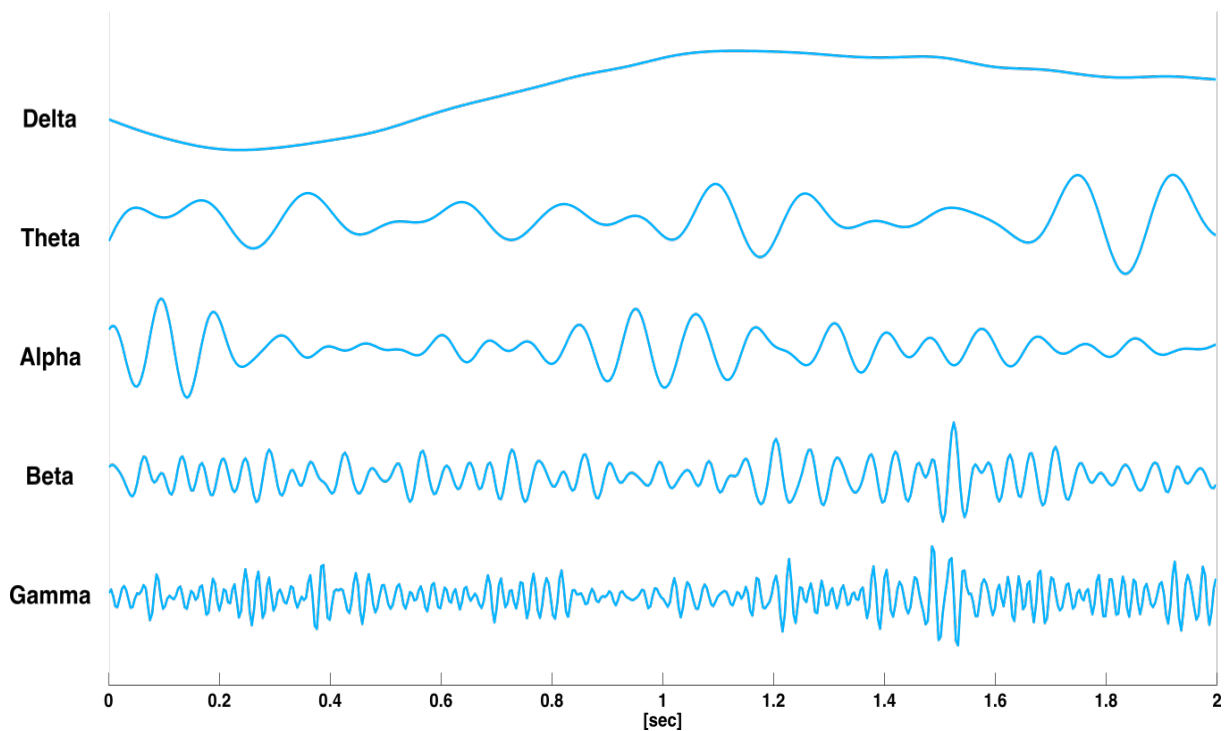
Fig. 2.1: Main EEG Rhythms. The electrical activity generated by neurons can be separated into frequency bands, which have been associated to different brain functions. In this way, by using spectral techniques, it is possible to independently analyze patterns within the different brain rhythms.

2.1.2 International 10-20 system

The way how sensors are positioned along the scalp for EEG acquisition is collectively referred as *electrode placement* [1]. A standard technique to cover the different regions of interest over the head is known as *International 10-20 system* [44], which owes its name to the proportional distance separating the electrodes between each other. To this end, nasion (e.g., top of the nose) and inion (e.g., most prominent point of the occipital bone) are taken as references, from which five aligned electrodes are placed at 10%, 20%, 20%, 20% and 10% of the distance separating both points. By using the same proportion, 16 more electrodes are placed to cover the area inside the diameter defined by the aforementioned references. This configuration leads to a total of 21 electrodes, which are represented in black in figure 2.2. The standard nomenclature that is used to identify electrodes by the major skull bones is indicated on each one of the sensors [45]. This nomenclature uses a letter for referring to the principal lobe where electrodes are located, and identifies all sensors either on the left or right sides by using, respectively, odd and even numbers, while electrodes along the midline are indicated by "z" ("z"=zero).

In addition, there are two other electrode placements corresponding to systematic extensions of the standard clinical 10-20 standard, namely, 10-10 and 10-5 placement systems [46]. The total of sensors comprised by the 10-10 system includes the 21 electrodes of the 10-20 placement together

with 53 additional electrodes, which are represented in gray in figure 2.2. Possible intermediate electrodes comprised in the 10-5 system are indicated by dots or white circles. The 68 white positions (named by extending the standard nomenclature), together with the 74 electrodes included in the 10-10 system, result in a subset of 142 electrodes that affords a more exhaustive and uniform coverage of the head [44]. All these techniques are widely, although not universally, used. In practice, electrodes montage varies across research groups, specially when a large amount of channels (> 64) is required, for which other placement systems have been developed in order to obtain a more regularly spaced sampling of scalp potentials [47].

2.2 Primary motor cortex

The cerebral cortex is an accessible area to electrode probes as well as to scalp recordings where the functionality of executive motor and communication behaviors take place, thus it represents a region of particular interest in BCI research [48].

The characteristics of the cortical surface provide convenient landmarks for the identification of particular regions within the brain, which are often used to guide implementation of intracortical electrodes. Back in 1870, Eduard Hitzig and Gustav Fritsch observed the movement of a dog's limb while the surface of its brain was being electrical stimulated on the opposite side [49]. This observation was crucial in many respects. On the one hand, it demonstrated that, as muscles, the brain is electrically excitable. And, on the other hand, by making the link between limb movement and the excitation of a particular region, the important interrogation of whether different brain areas, more precisely the cerebral cortex, were engaged to different functions was confirmed. The relation between particular regions and the control of particular body parts was subsequently coined as *somatotopic organization*. Based on their experiments, Hitzig and Fritsch ultimately described the area located in the frontal lobe of the cerebral cortex that we now know as the *primary motor cortex*.

Nowadays, in order to identify *eloquent cortex* (i.e., areas where damage will result in paralysis, loss of sensation, or linguistic impairment), neurosurgeons consistently map the brains of awake human patients undergoing surgical procedures for treatment of severe epilepsy or tumor resection by using electrical stimulation. These techniques were first applied by Wilder Penfield, whose work led to the now well-known map of the *motor homunculus* [50], [51]. This map represents the areas of the motor cortex associated with different motor functions. It is a distorted image of the body in which parts requiring more precisely graded control (e.g., the hand, and more specifically the thumb) have disproportionally large representations.

Figure 2.3 shows an adaptation of Penfield's homunculus, where it is possible to see the distribution of the different body parts along their contralateral area over the primary motor cortex in accordance with electrodes C3, C_z, and C4 of the international 10-20 system. Legs and feet are represented within the medial wall of the cerebrum; the trunk, upper arm, and hand are found consecutively and laterally along the hemisphere, with the face located at the most
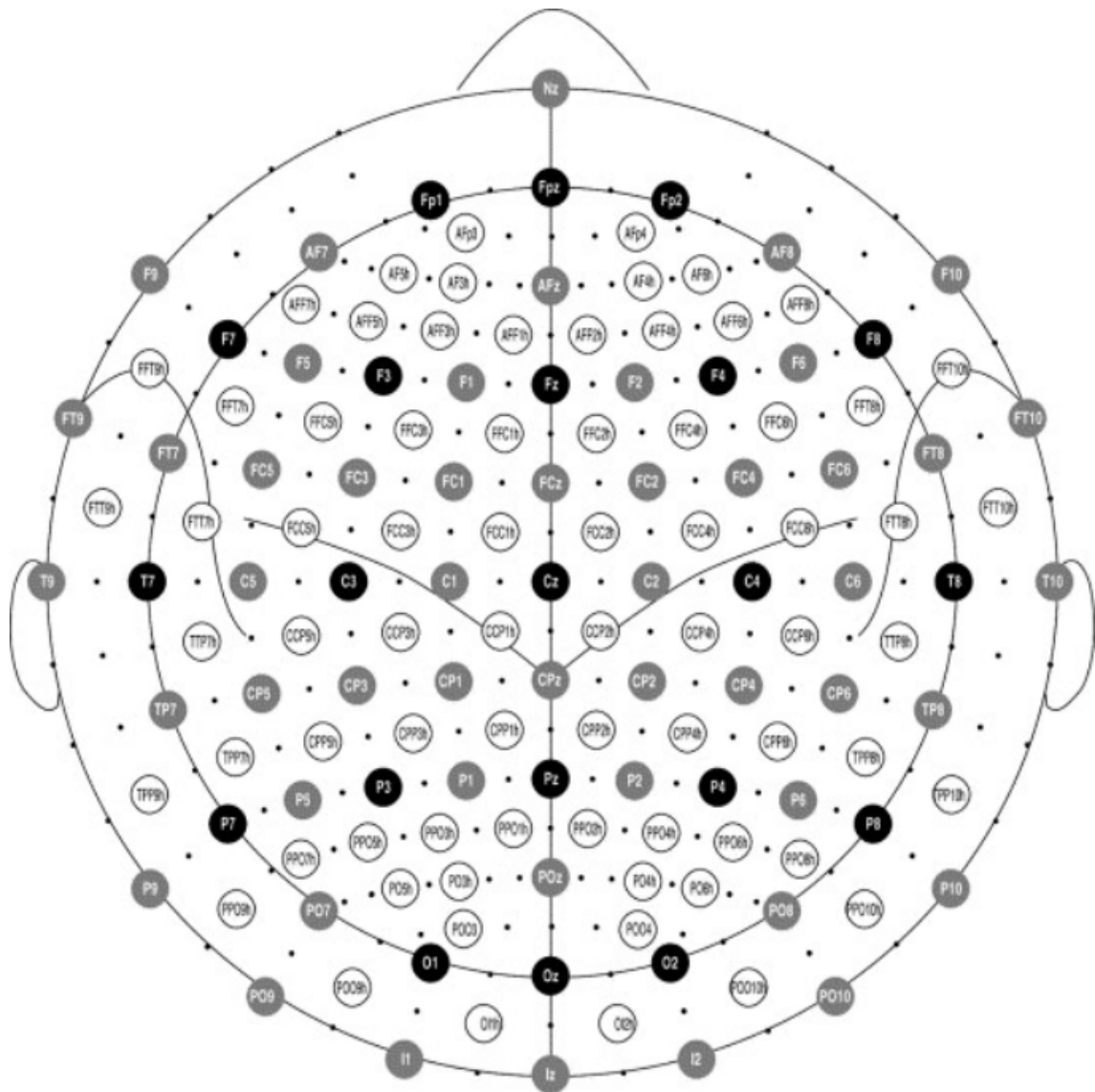
Fig. 2.2: The standard 10-20, 10-10 and 10-5 electrode montages. The 10-20 montage is indicated by the 21 electrodes represented as black circles. The 10-10 montage, which has a total of 74 electrodes, includes the 21 electrodes of the 10-20 montage plus 53 additional ones indicated by gray circles. The 10-5 montage comprises the 74 electrodes of the 10-10 system together with 68 additional ones represented by dots and white circles, which gives a total of 142 electrodes that provide a more exhaustive and uniform coverage of the head. The nomenclature for the standard 10-20 and 10-10 and 10-5 extensions is indicated. Note that electrodes on the right side have even numbers and electrodes on the left side have odd numbers, while electrodes along the midline are indicated by "z" ("z"=zero) [46].

lateral. Although neighboring body parts are typically represented by the homunculus within neighboring areas, they are spatially distorted given that some of them require a more complex control, which demand a bigger surface. For instance, control of the several muscles involved in facial or hand movement is much more complex than it is for the much larger biceps muscle that

flexes the elbow. Consequently, a greater amount of cortical area is engaged in the control of the face or hand than to the upper arm. These general principles of somatotopic organization can be observed in sensory as well as motor areas of the cortex.
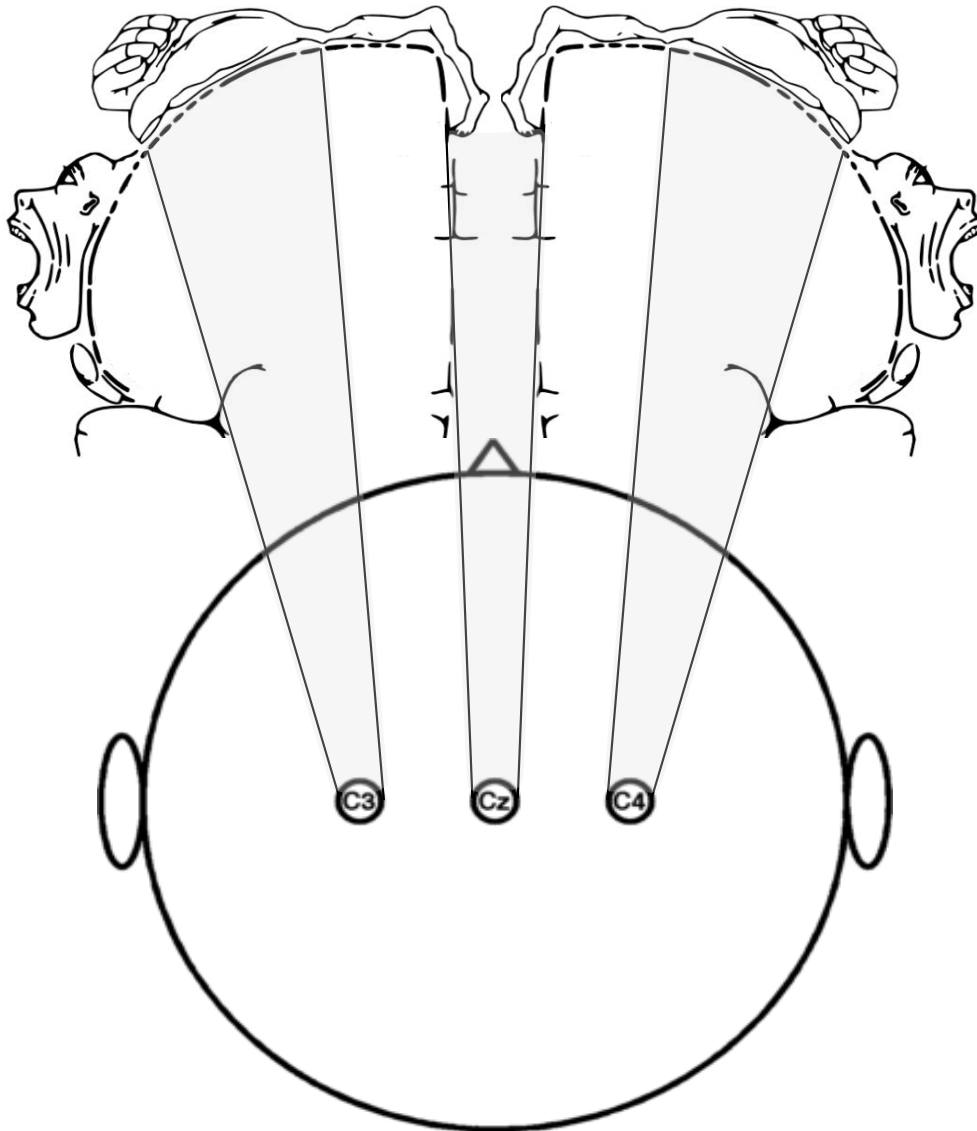
Fig. 2.3: The motor homunculus by Wilder Penfield represents the somatotopic organization that maps the areas within the motor cortex associated with different motor functions. Electrodes C3, C_z, and C4 have been included to show the location of the different body parts with respect to the international 10-20 system.

2.3 Sensorimotor rhythms

Sensorimotor rhythms (SMRs) are oscillations in the electric or magnetic fields registered over sensorimotor cortices (i.e., posterior frontal and anterior parietal areas) [52], which are mostly

found between three main frequency bands: mu (8 – 12 Hz), beta (14 – 30 Hz) and gamma ($>$ 30 Hz). In the case of EEG, recordings are mostly limited to mu, beta and lower-frequency gamma activity, whereas ECoG and MEG signals can be modulated within higher frequencies [1].

2.3.1 Event-related desynchronization

The transitory amplitude decreasing of oscillations in the alpha and beta bands taking place in localized regions and in direct relation to an event is known as event-related desynchronization (ERD) [53]. This phenomenon can be observed not only during, but also before, sensory and cognitive processing and motor behavior of the awake brain [54]–[57]. Therefore, ERD is not only an electrophysiological correlate of cortical activation related to stimulus processing or motor output, but is also characteristic for cortical areas or neural structures preparing to process sensory information or ready and prepared to execute a motor command [5].

2.3.2 Event-related synchronization

The opposite phenomenon is known as event-related synchronization (ERS), which consists of a localized increase of power in the alpha and beta bands generated by bursting activity [58]. Both ERS and ERD have a specific topographical distribution depending upon the state of the brain, and they can be either externally (e.g., by stimuli) or internally (e.g., by voluntary behaviour) triggered; which in terms of the BCI terminology correspond, respectively, with synchronous and asynchronous events.

Both cortical and subcortical structures take part as extensive networks responsible for the SMRs observed within the cortex. The use of depth electrodes has revealed mu and beta activity along the thalamus, subthalamic nucleus, and penduculopontine area [59]–[61]. The relationships among these areas are complex. For instance, during movement, mu ERD occurs within the motor cortex and mu ERS occurs in the subthalamic nucleus, whereas beta ERD occurs uniformly in motor cortex, thalamus, and the subthalamic nucleus [60].

2.3.3 Lateralization

As mentioned, ERD corresponds with an activity pattern of sensorimotor rhythms occurring during real or imagined movements [5]. ERD% lateralization, i.e. the difference between the relative power changes of the contra- and ipsilateral hemispheres, constitutes the basis of the majority of binary MI-based BCIs [5], [62]–[68].

Previous studies also reported a significant ERD over the contralateral side and a significant ERS over the ipsilateral side of the brain for planned and terminated movements, respectively. For example, alpha and beta ERD begin contralaterally about 1.5 - 2 s before self-paced movement, becoming bilateral during movement execution [11], [69]–[74]. The significant ipsilateral ERS was found to be characteristic of the beta band for both after physical and imagined termination of motor tasks, which has been referred as post-movement rebound [11], [12], [75]–[77].

2.3.4 Spatial mapping of ERD/ERS

EEG multichannel recordings are frequently recorded against a common reference electrode, thus data are reference-dependent. In order to convert reference-dependent raw data into reference-free data, different techniques are available, such as Common Average Reference (CAR), Hjorth Laplacian Reference (HR), and Local Average Reference (LR) [3], [78], [79]. Reference-free topographic maps typically provide better defined patterns than referential representations and are especially recommended for sensorimotor rhythms analysis [11]. For spatial mapping of ERD/ERS it is possible to use a spherical head model together with an interpolation method (e.g., cubic interpolation).

Figure 2.4 presents a series of topographic maps showing the sensorimotor rhythms over the mu band induced by a real left hand movement (fig. 2.4-A) and an imagined left hand movement (fig. 2.4-B), in comparison with the activity induced by a real right hand movement (fig. 2.4-C) and an imagined right hand movement (fig. 2.4-D). EEG signals belong to a right-handed male subject that has participated in the development of the database that will be described along the following sections. In all maps it is possible to observe the ERD phenomena characteristic of the execution of a motor task, which is represented in blue colors associated to the minimum values. For both left and right hand motor tasks, patterns are more enhanced in the case of real movements, which in comparison to the motor imageries, present darker tonalities and more defined contours around the associated electrodes over their contralateral side (C4 for left hand and C3 for right hand, see fig. 2.3 in regard of the lateralization of the somatotopically organization).

Figure 2.5 presents a series of topographic maps showing the sensorimotor rhythms over the beta band induced after the termination of a real left hand movement (fig. 2.5-A) and an imagined left hand movement (fig. 2.5-B), in comparison with the activity induced after the termination of a real right hand movement (fig. 2.5-C) and an imagined right hand movement (fig. 2.5-D). EEG signals belong to the same subject considered in the previous example. In all maps it is possible to observe the ERS phenomena characteristic of the termination of a motor task (e.g., post-movement rebound), which is represented in red colors associated to the maximum values. As in the previous example, ERS patterns are observed around the associated electrodes over the contralateral side (C4 for left hand and C3 for right hand, see fig. 2.3).

2.4 Event-related potentials

Different events can induce time-locked changes in the activity of neural populations, which are typically known as Event-Related Potentials (ERPs), from which the most remarkable correspond to sensory stimuli responses [11]. Averaging techniques are generally applied in order to enhance the morphology of ERPs. The underlying premise is that the patterns of interest have a relatively constant time-delay to the stimulus, while the ongoing activity behaves as additive noise, thus the averaging process improves the signal-to-noise ratio. However, there are some omissions that do not hold to this assumption; on the one hand, it is known that Evoked Potentials (EPs) can
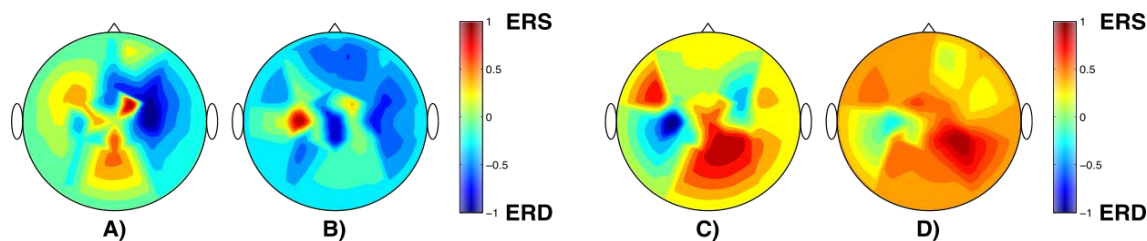
Fig. 2.4: Topographic maps during the course of motor tasks within the mu band. A) presents the brain activity generated during a real left hand movement execution, whereas B) shows the activity generated during a left hand imagined movement. In both cases it is possible to observe the characteristic ERD patterns that are induced in the contralateral hemisphere (around C4, see fig. 2.3), although these are more enhanced in the case of the real movement. C) and D) show, respectively, the real movement execution and the imagined movement of the right hand. In the same way, it is possible to observe the lateralization of the ERD patterns associated to the motor tasks (around C3, see fig. 2.3). Spatial mapping was carried out by using a cubic interpolation method [80]. Note that each pair of maps was independently normalized.
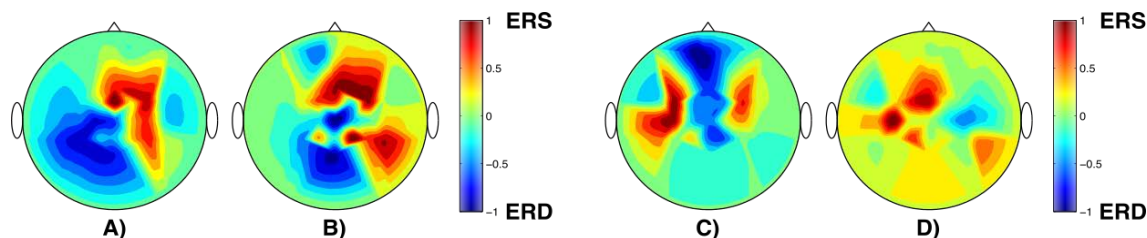
Fig. 2.5: Topographic maps after the termination of motor tasks within the beta band. A) presents the brain activity generated after a real left hand movement execution, whereas B) shows the activity generated after a left hand imagined movement. In both cases it is possible to observe the characteristic ERS patterns (e.g., post-movement rebound) that are induced in the contralateral hemisphere (around C4, see fig. 2.3). C) and D) show, respectively, the termination of a real movement execution and the termination of an imagined movement of the right hand. In the same way, it is possible to observe the lateralization of the ERS patterns associated to the motor tasks (around C3, see fig. 2.3). Spatial mapping was carried out by using a cubic interpolation method [80]. Note that each pair of maps was independently normalized.

be considered to result from a reorganization of the phases of the ongoing EEG signals [81]. Furthermore, it was also demonstrated that visual stimuli can reduce the amplitude of the ongoing EEG activity [82], showing that the model assuming that an ERP can be represented by a signal added to uncorrelated noise does not hold in general. And on the other hand, considering that the changes evoked by certain events that block or desynchronize the ongoing alpha activity are time-locked to the event but not phase-locked, they cannot be extracted by a simple linear method, such as averaging, but may be detected by frequency analysis. In summary, event-related phenomena cause frequency specific changes of the EEG activity which, in general terms, consist either of decreases or of increases of power in particular frequency bands. This may be considered to be caused, respectively, by a decrease (ERD) or an increase (ERS) in synchrony of the underlying neural populations [3], [53], [70].

Contrary to the classical ERPs, which can be characterized as a series of transient post-synaptic responses of main pyramidal neurons responding to a particular stimulus, ERD/ERS phenomena can be considered as the result of changes in one or more variables that control oscillations in neural networks. The properties of EEG oscillations are determined by [11], [83], [84]:

1. the intrinsic membrane properties of the neurons and the dynamics of synaptic processes;

2. the strength and extent of the interconnections between the network elements, most often formed by feedback loops. Different kinds of feedback loops can be distinguished: involving thalamo-cortical or cortico-cortical either at short or at long distances;

3. the modulating influences from general or local neuro-transmitter systems.

2.5 Time course of ERD/ERS

Typically, ERD/ERS measurements embraced in particular frequency bands are represented relative to the power of the same EEG channels recorded during a baseline period a few seconds before the event occurs, as a percentage magnitude. Event related changes in ongoing EEG need time to thrive and to recover, especially when slow alpha band rhythms are involved, therefore the interval between two consecutive events must last at least some seconds. For voluntary limb movement studies, it is recommended to use an inter-interval of at least 10 s and, in the case of a foreperiod reaction time task, this interval must be even longer.

The classical method to compute the time course of ERD/ERS includes the following steps [11]:

1. bandpass filtering of all event-related trials within the frequency band of interest;

2. squaring of the amplitude samples to obtain power samples;

3. averaging of power samples across all trials;

4. averaging over time samples to smooth data and reduce variability.

To obtain ERD/ERS values in terms of percentage, the power within the frequency band of interest following the event is given by x, whereas that of the preceding baseline or reference period is given by BL. ERD/ERS values are computed as the percentage of power decrease or increase, respectively, according to the following expression:

$$\text{ERD/ERS\%} = \frac{\overline{x^2} - \overline{BL^2}}{\overline{BL^2}} \times 100, \qquad (2.1)$$

where $\overline{x^2}$ is the mean of the squared EEG signal, $\overline{BL^2}$ is the mean of a baseline sliding window taken at the beginning of the corresponding trial, and ERD/ERS% is the percentage of the oscillatory power estimated at each step of the sliding window. Oscillations greater than the baseline during the MI task produce positive percentage values that are assumed as a modulation of ERS, whereas

the negative values generated by smaller oscillations are considered as a modulation of ERD.

Figure 2.6 shows a detailed example of the time course of ERD/ERS% computation according to equation 2.1. For which 80 trials of right hand motor imagery recorded over electrode C3 have been considered. EEG signals belong to a right-handed female subject that has participated in the development of the database that will be described along the following sections. Figure 2.6-A shows the raw EEG signals. Note that, because of space limitations, only 3 trials are displayed but that the process includes the entire set. The motor imagery starts at second "0" and lasts along 6 seconds, which is indicated by a blue segment. As mentioned, the first step consists of the band-pass filtering of the signals which, in this case is carried out throughout a 5th order Butterworth filter within the mu band in order to enhance the characteristic ERD pattern associated to this band-with (see figure 2.6-B). Subsequently, the band-pass filtered signals are independently squared across samples (figure 2.6-C), where the power decrease is already pronounced along the motor imagery period. Figure 2.6-D presents the averaging across trials, where a 2-second bar has been drawn 2 seconds before the initiation of the motor task to indicate the baseline segment that was considered in order to smooth the signal according to equation 2.1 by using a 125 ms shifting step. Finally, the resulting time course of ERD/ERS is shown in figure 2.6-E. Negative values reveal a power decrease with respect to the baseline (reference) segment, which correspond to an ERD modulation. From the figure it is possible to observe that this modulation (highlighted in blue) takes place approximately 0.5 s before the motor imagery starts, and that it lasts for some seconds following its termination, after which it begins to recover slowly to the reference line and going beyond to positive values (highlighted in red), which correspond to the ERS modulation characteristic of the post-movement rebound.

2.6 Combined movements

In contrast to simple motor tasks, combined imagined movements consider the use of two or more body parts at the same time. In this way it is possible to dramatically increase the number of commands afforded by a BCI system. However, as presented along the introduction of this document, most of the research has been centred on the EEG patterns induced by simple motor imageries, and less work has been reported regarding the study of compound tasks. In this sense, to our best knowledge, the question whether the EEG activity induced by combined imagined movements can be characterized as the superposition of the activity that is independently generated by the sources associated to the body parts engaged in the task. To first approach this question, we have a created a 4-class database comprising the rest condition and motor imageries of left hand, right hand and both hands together. With which a statistical analysis was performed in order to determine whether or not the activity in the associated areas (around electrodes C3 and C4, see fig. 2.3) behaves the same in the case of the combined task with respect to each one of the simple motor imageries.
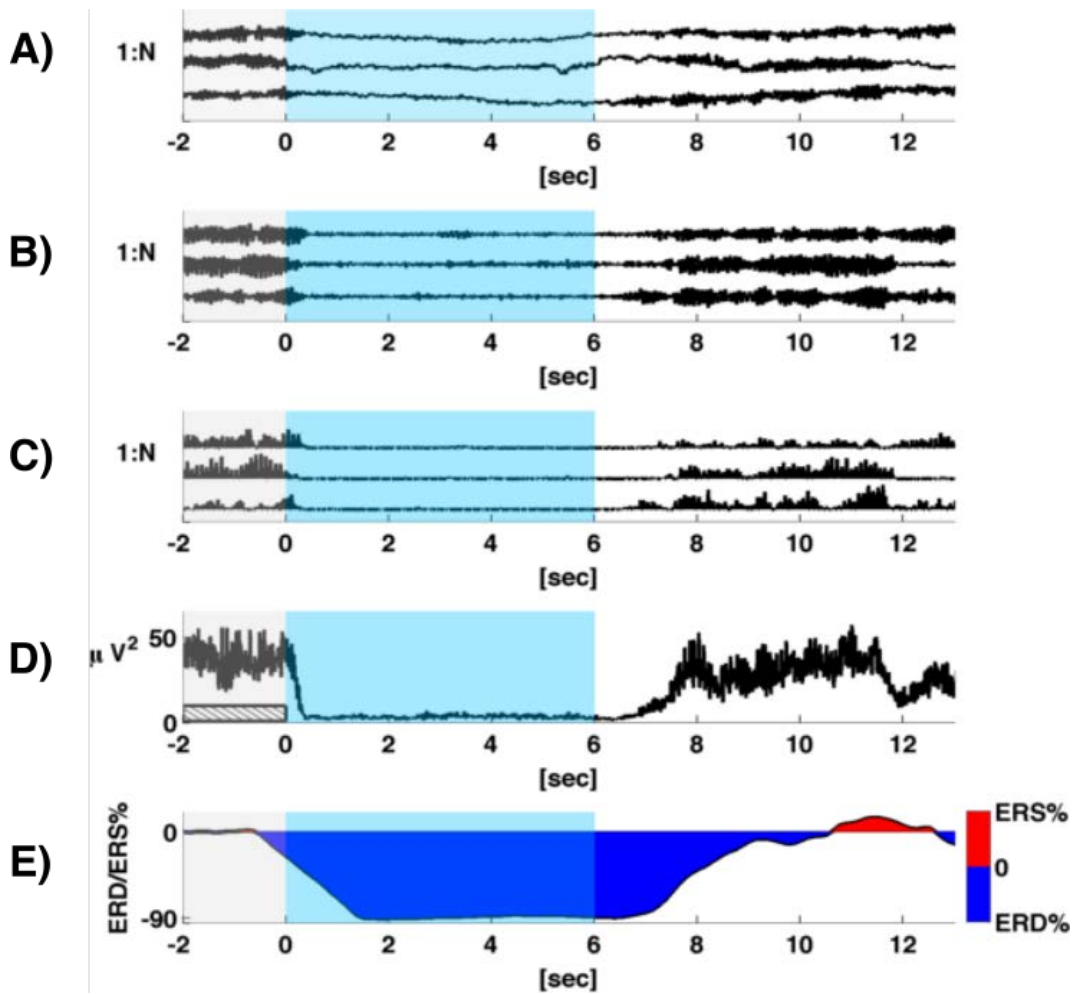
Fig. 2.6: Processing method to observe the time course of ERD/ERS% according to equation 2.1. EEG segments correspond to right hand motor imageries recorded on electrode C3 (A), (B) shows the resulting band-pass filtered signals within the mu band by using a 5th order Butterworth filter, (C) illustrates the squaring of samples, (D) presents the averaging across trials, and (E) depicts the resulting time cours of ERD/ERS highlighted, respectively, in blue and red colors.

2.6.1 4-class database

EEG signals were recorded through the openViBE [1] [85] platform from five right-handed healthy subjects at 256 Hz using a commercial REFA amplifier developed by TMS International™.

Subjects were asked to perform three different motor imageries (left hand, right hand, and both hands) together with rest condition, while seated in a comfortable chair with the arms at their sides in front of a computer screen showing the task cue to be performed. The whole session consisted of 4 runs containing 20 trials per task for a total of 80 trials per class. All experiments were carried out with the consent agreement of each participant and following the statements of the WMA declaration of Helsinki on ethical principles for medical research involving human

---

1. `http://openvibe.inria.fr`

subjects.

2.6.1.1 Paradigm and time scheme

The EEG cap was fitted with 26 electrodes, namely, Fp1; Fp$_z$; Fp2; F$_z$; FC5; FC3; FC1; FC$_z$; FC2; FC4; FC6; C5; C3; C1; C$_z$; C2; C4; C6; CP5; CP3;CP1; CP$_z$; CP2; CP4; CP6, and P$_z$, re-referenced with respect to the common average reference across all channels and located over the international 10-10 system positions to cover the primary sensorimotor cortex (see fig. 2.7).
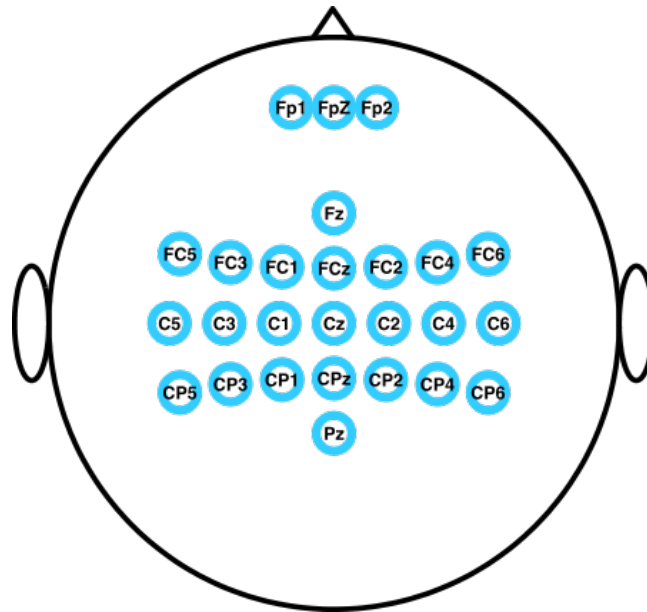
Fig. 2.7: Distribution of the 26 electrodes considered for the 4-class database development. The EEG cap was fitted with 26 electrodes re-referenced with respect to the common average reference across all channels and located over the extended international 10-20 system positions to cover the primary sensorimotor cortex.

Two panels were simultaneously displayed on the screen, which were associated from left to right, to the left hand and right hand. Each trial was randomly presented and lasted for 12 seconds, starting at second 0 with a cross at the center of each panel and an overlaid arrow indicating for the next 6 seconds the task to be performed: an arrow pointing to the left side on the left panel for left hand, an arrow pointing to the right side on the right panel for right hand, and the simultaneous combination of the two of them for the corresponding compound motor imageries (see fig. 2.8). The rest condition was also considered and it was indicated by the absence of arrows. After second 6, the task cue disappeared and the crosses were remaining for the next 6 seconds indicating the pause period before the next trial started.

2.6.2 ERD/ERS% analysis

Each trial was considered 2 seconds before the task cue appeared on the screen and 6 seconds after it disappeared, so that each segment corresponds to a 14s-length signal over 26 electrodes.
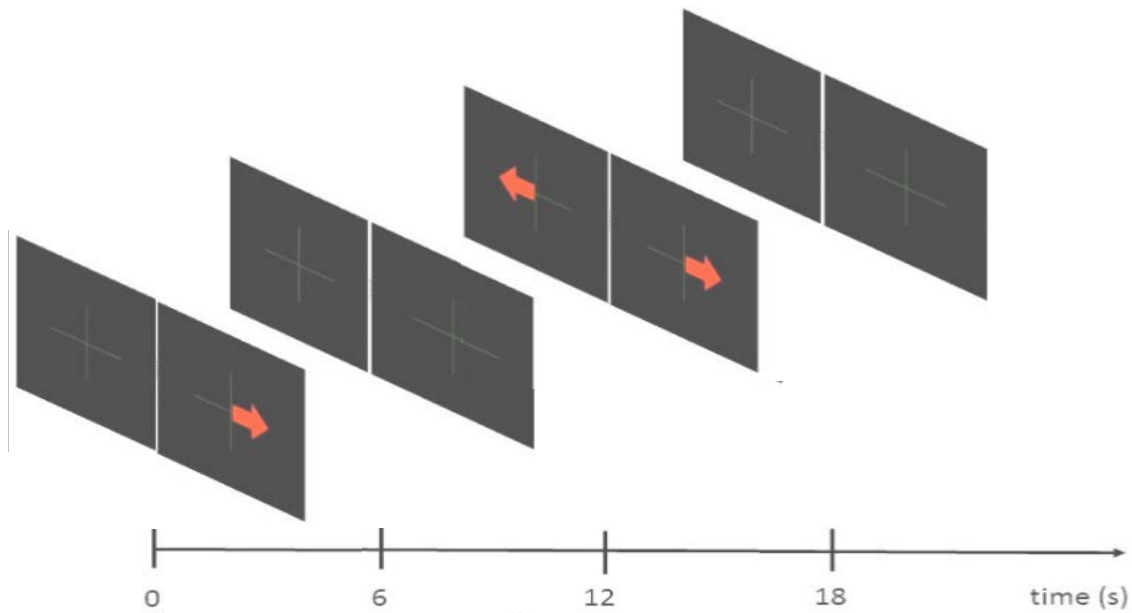
Fig. 2.8: Time scheme for the 4-class database. Each trial was randomly presented and lasted for 12 seconds. During the first 6 seconds, users were asked to perform the motor imagery indicated by the task cue, which was launched throughout two simultaneously displayed panels associated, from left to right, to the left hand and right hand. The use of each body part was indicated by the presence of arrows: an arrow pointing to the left side on the left panel for left hand, an arrow pointing to the right side on the right panel for right hand, and the simultaneous combination of the two of them for the corresponding compound task. The rest condition was also considered and it was specified by the absence of arrows. After second 6, the task cue disappeared and the crosses were remaining for the next 6 seconds indicating the pause period before the next trial started.

All segments were filtered using a 5th-order Butterworth filter within the mu and beta bands, in conformance to the nature of SMRs described in section 2.5 in page 24.

The time course of ERD/ERS% of each segment was computed by squaring its samples and smoothing the resulting signal through a 2-second-sliding window with a 125 ms shifting step according to equation 2.1. Figure 2.9 presents a topographic representation of the grand average across subjects of ERD/ERS% values, which were computed in terms of equation 2.1 over all electrodes within the mu range during the 6-second period of the four motor tasks. Over the contralateral hemisphere it can be observed that, during simple motor imageries, ERD% values are similar to those generated in the corresponding side throughout the combined task, which considering the entire set of electrodes, present the minimum values around C3 and C4 during the course of the three motor tasks. In the case of left hand motor imageries, the minimum ERD% values are induced around electrode C4, contrastingly to the activity observed during right hand motor imageries, which presents the lowest ERD around electrode C3, whereas for the combined task, the minimum power is found simultaneously over electrodes C3 and C4 (see fig. 2.3). On the other hand, over the ipsilateral hemisphere, simple motor imageries present similar ERS%

values to the observed ones during the rest condition.
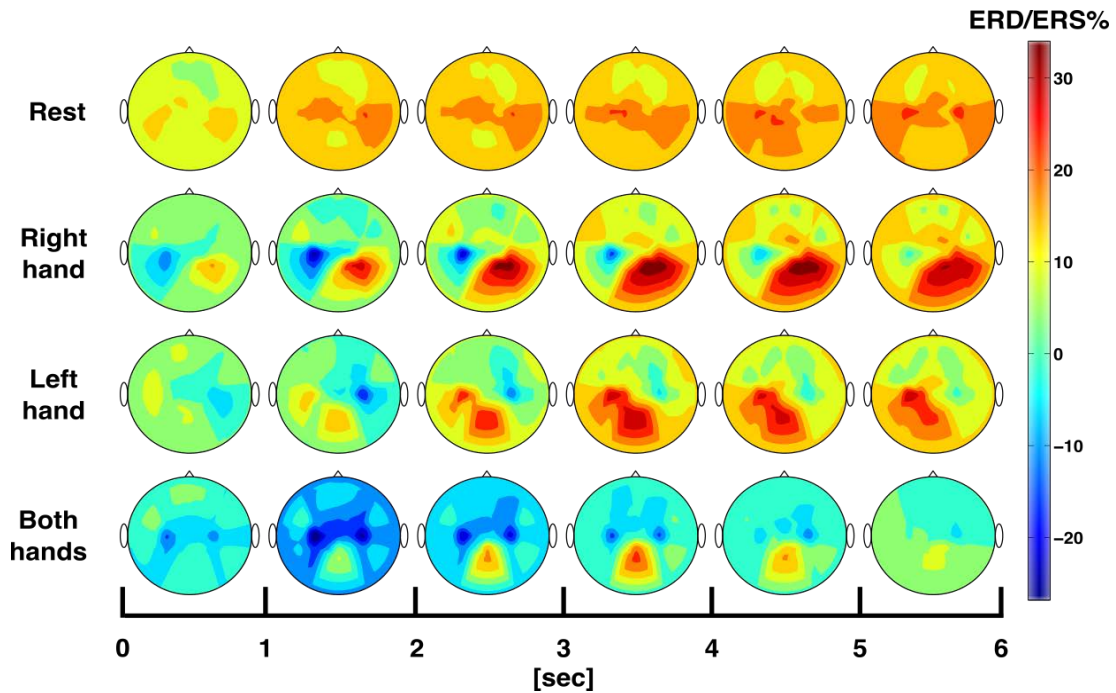
Fig. 2.9: ERD/ERS% topographic representations within the mu range during the 6-second period of the four motor tasks. The spatial mapping corresponds to the grand average across subjects (n=5) of the time course values computed in terms of equation 2.1 over all electrodes by using a cubic interpolation method [80].

To take a clearer look at the activity generated over the hand-related areas, the grand average across subjects of the time course of ERD/ERS within the mu range is presented in figure 2.10 for all motor tasks on electrode C3 (fig. 2.10-A) and electrode C4 (fig. 2.10-B). It can be observed that simple motor imageries behave, over their contralateral side, in the same way that the combined task. On electrode C3 (left hemisphere) the time course of right hand motor imageries represented by the red trace has the same morphology as the black one representing the combined movement, both showing the lowest ERD with the minimum value around second 2. In the same way, when looking at electrode C4 (right hemisphere), the blue trace representing the time course of left hand motor imageries shows slightly the same behaviour as the combined task represented by the black trace, in this case both elements present also the lowest ERD values and reach their minimum around second 2.

Furthermore, in the ipsilateral hemisphere, simple motor imageries produce similar ERS% values to the ones estimated on both sides for the rest condition, which is represented by the green trace.

The comparison between the rest condition together with the three motor imageries once the motor tasks were finished is shown in figure 2.11. Which shows, during the 6 seconds following the motor tasks, the corresponding spatial maps computed over all electrodes of the grand average across subjects of the ERD/ERS% values estimated in terms of equation 2.1 within the beta range.

29
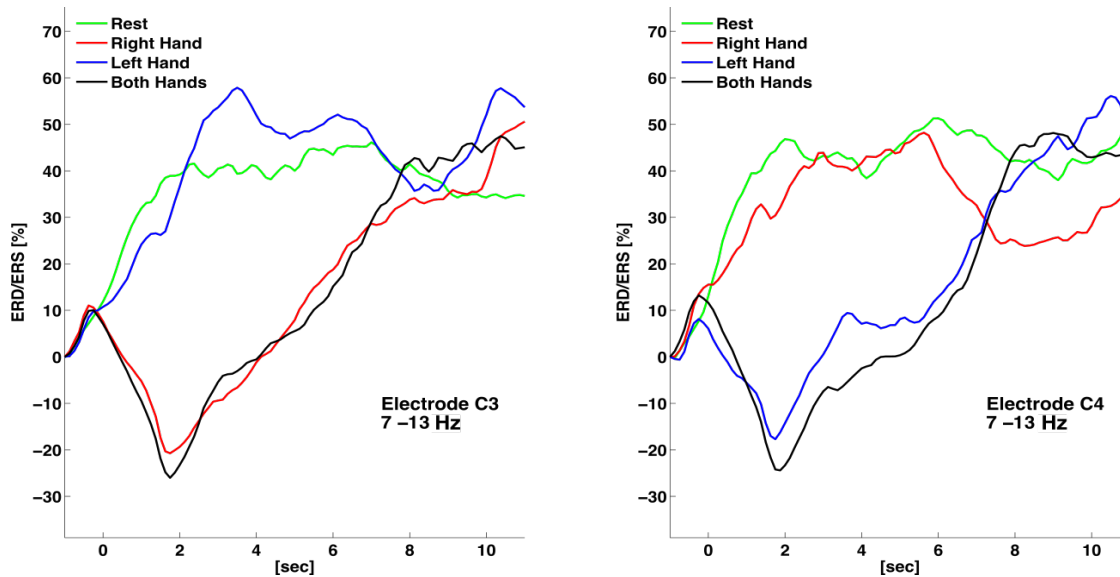
Fig. 2.10: ERD/ERS% time series within the mu range for all motor tasks over electrodes C3 and C4. ERD/ERS values were estimated according to equation 2.1 over the grand average accross subjects (n=5). The oscillatory power on electrodes C3 and C4 is presented, respectively, in (A) and (B).

It can be observed that the activity is distributed in a similar way as during the motor tasks (i.e., around electrodes C3 and C4, see fig. 2.9), but with much higher values; which indicates that the corresponding regions present an ERS modulation characteristic of the termination of motor tasks (i.e., post-movement rebound). In the case of left hand motor imageries, the maximum ERS% values are induced around electrode C4, contrarily to the activity observed after right hand motor imageries, which presents the highest ERS around electrode C3, whereas for the combined task, the maximum power is found simultaneously over electrodes C3 and C4. On the other hand, over the ipsilateral hemisphere, simple motor imageries present similar ERS% values to the observed ones during the rest condition.

The grand average across subjects of the time course of ERD/ERS within the beta band for all motor tasks over electrodes C3 and C4 is shown, respectively, in figure 2.12-A and figure 2.12-B. Similar to figure 2.10, though with smaller magnitude, there are significant ERD% values over the contralateral side during the course of simple motor imageries, which present a comparable morphology to the one induced by the combined task on both sides. However, it worth noting that, once the motor tasks have finished, traces go back to positive values much faster than in the mu band, phenomena that has been already reported by [10]. Furthermore, the ERS% values observed after the simple motor tasks termination, which correspond to the so-called post-movement rebound, present the same lateralization that the ERD% values observed during the course of motor imageries, so that in the case of left hand motor imageries, the highest ERS% values take place over electrode C4 (right hemisphere), whereas for the right hand motor imageries, the highest ERS% values are induced in electrode C3 (left hemisphere). In the case
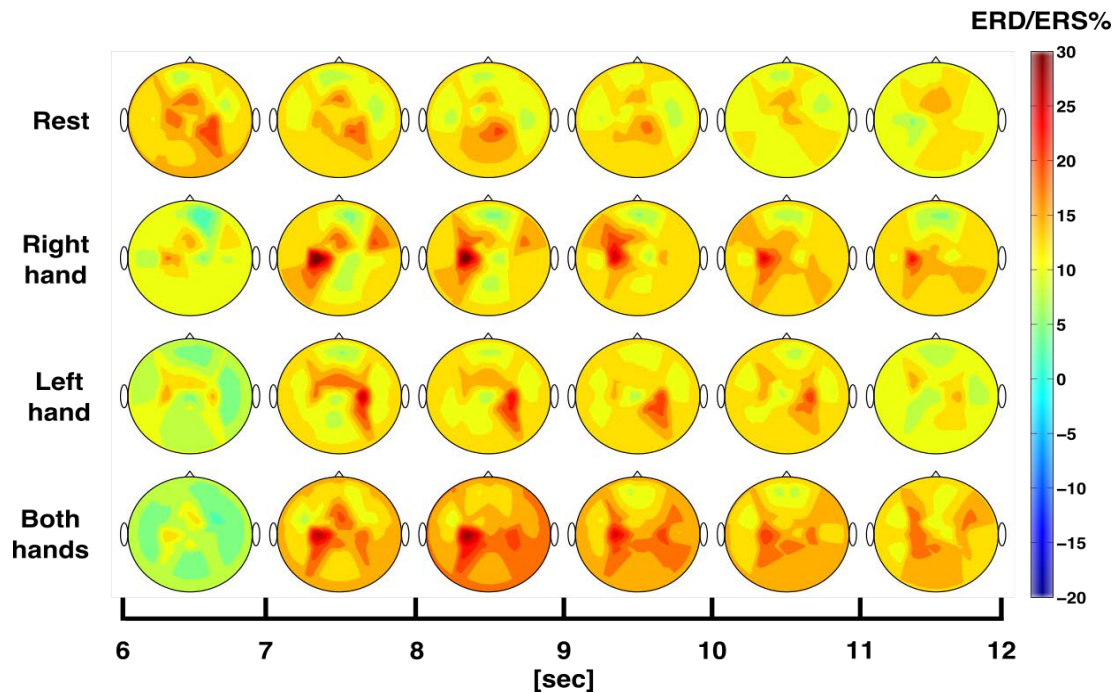
Fig. 2.11: ERD/ERS% topographic representations within the beta range during the 6-second period after the termination of the four motor tasks. The spatial mapping corresponds to the grand average across subjects (n=5) of the time course values computed in terms of equation 2.1 over all electrodes by using a cubic interpolation method [80].

of the combined task, the post-movement rebound is appreciable on both sides although, in comparison with the one generated by simple motor imageries, it presents significantly smaller values. On the other hand, over the ipsilateral hemisphere after simple motor imageries, there is no post-movement rebound and the time course of the corresponding traces looks similar to the one generated by the rest condition.

2.6.3 Statistical analysis

From figure 2.10 and figure 2.12 it can be observed that the lowest ERD% values are reached within the interval [1-2 s]. For this reason, in order to better evaluate the differences between conditions that take place during the course of the considered motor tasks, all segments were analyzed only during such a period. Furthermore, since the ERD% values characteristic of a motor task execution are well observed over both mu and beta bands, all trials were band-pass filtered in the range from 8 to 25 Hz (e.g., mu-band + beta-band) by using a 5th order Butterworth filter, within which the power spectrum was computed across all channels. Subsequently, the aforementioned range was scanned by taking subintervals of 2 Hz each to compute, in terms of R^2, the correlation between the power spectrum magnitude over all pairs of conditions at each electrode [4].

High R^2 values indicate that there is a significant difference between the considered motor
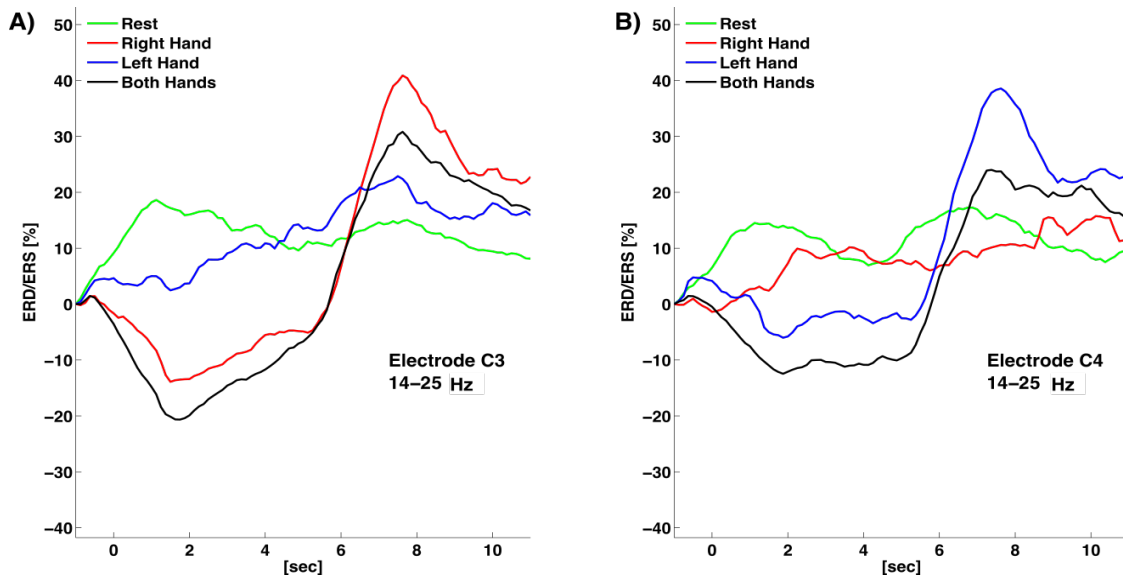
Fig. 2.12: ERD/ERS% time series within the beta range for all motor tasks over electrodes C3 and C4. ERD/ERS values were estimated according to equation 2.1 over the grand average across subjects (n=5). The oscillatory power on electrodes C3 and C4 is presented, respectively, in (A) and (B).

tasks at the given frequency and over the corresponding positions. On the contrary, small R^2 values show that the conditions under analysis do not present a significant difference, and therefore are similar between each other [21].

Figure 2.13 shows the resulting R^2 values generated for each comparison among all classes. Figure 2.13-A presents the comparison between left hand motor imageries and the rest condition; it can be observed that the highest R^2 values are found at 12 Hz over the right hemisphere around electrode C4. This is due to the fact that, for the rest condition, there are ERS% values in both hemispheres, whereas for left hand motor imageries there are ERS% values only in the ipsilateral side, since the contralateral presents ERD%. Figure 2.13-B shows the comparison between left hand motor imageries and the combined task. In this case the highest R^2 values are found in the left side around electrode C3, since for both hands motor imageries there are ERD% values in both hemispheres, whereas for the left hand task there are ERD% values only in the contralateral side, given that the ipsilateral hemisphere presents ERS%. Figure 2.13-C and figure 2.13-D present, respectively, the comparison between right hand motor imageries and rest condition, and the comparison between right hand motor imageries and both hands motor task. The situation is similar as the one described for the left hand motor task but with the highest R^2 values found in the other way around, since the ERD/ERS% values occur in the same manner but with respect to the right hemisphere (electrode C3). Figure 2.13-E shows the comparison between left hand and right hand motor imageries. In this case, the highest R^2 values are also at 12 Hz and simultaneously distributed around electrodes C3 and C4. This is because the brain activity within both hemispheres presents an opposite behavior during the course of these two

types of motor imageries and, therefore, there is always a significant difference between each other. In the same way, the comparison between both hands motor imageries and the rest condition presented in figure 2.13-F, shows high R^2 values in both hemispheres; since the both hands motor task generates ERD% values in both sides, whereas the rest condition produces ERS%.
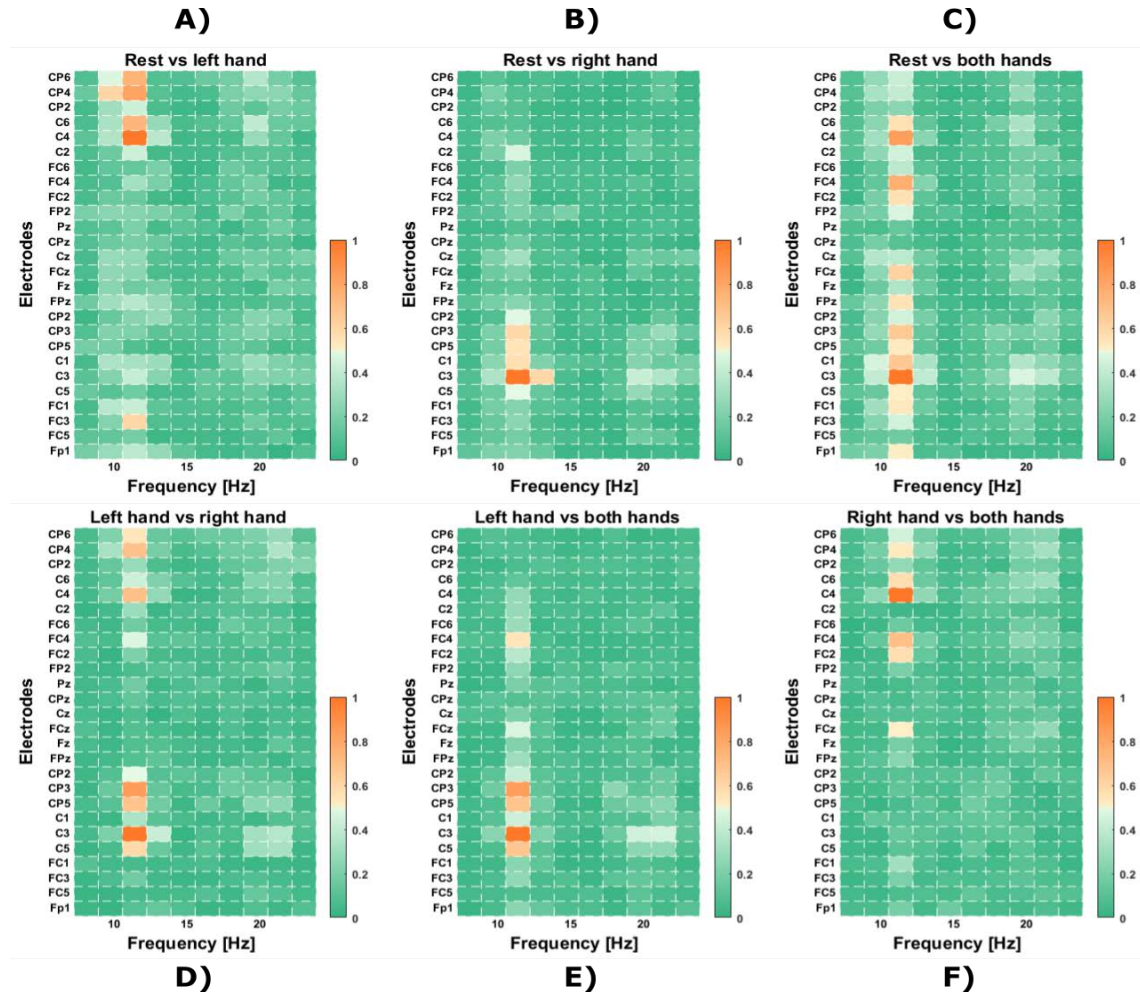
Fig. 2.13: Grand average across subjects (n=5) of correlation between the considered conditions over each electrode in terms of R^2. The different comparisons correspond to (A) left hand MIs and rest condition; (B) left hand MIs and both hands MIs; (C) right hand MIs and rest condition; (D) right hand MIs and both hands MIs; (E) left hand MIs and right hand MIs and F) both hands MIs and rest condition. High R^2 values indicate that there is a significant difference between the corresponding MIs at the given frequency over the corresponding electrode positions.

From figure 2.13 it can be observed that, for all comparisons, the most significant R^2 values occur around 12 Hz. Therefore, in order to visualize the correlation between electrodes in a more intuitive way, R^2 values within this frequency were used to compute the topographic maps that are shown in figure 2.14, where the comparisons for each pair of conditions are presented.

Along this section, it has been reported that during the course of the considered motor imageries, brain activity has important components around 12 Hz, and that it is distributed
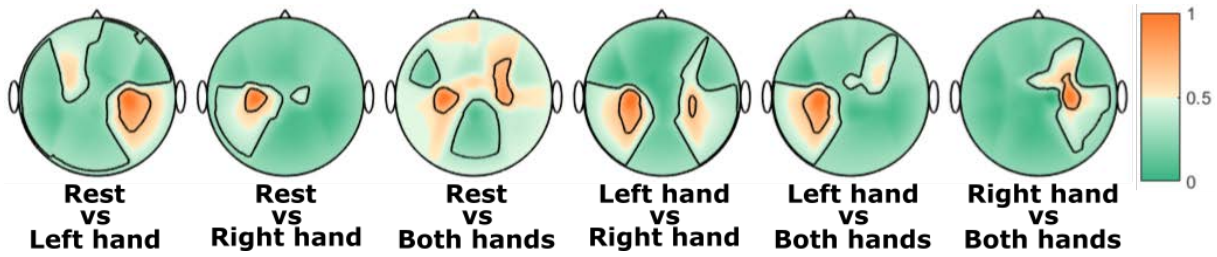
Fig. 2.14: Topographic representations of the grand average across subjects (n=5) of the correlation values between the considered conditions in terms of R^2.

over the primary sensorimotor cortex according to the body part that is engaged at each task, especially around electrodes C3 and C4 (see fig. 2.3). With this in mind, it is possible to estimate the level of significance between the considered conditions. To this end, the power spectrum of the one-second segment (i.e., from 1 to 2 s) of each filtered trial within the frequency range from 8 to 25 Hz over electrodes C3 and C4 was computed for each subject. Subsequently, all values at 12 Hz belonging to the same condition at each electrode were gathered together and, an unpaired t-test was applied to determine with a 5% level of significance, whether or not two different conditions present comparable activity over the left (electrode C3) and right (electrode C4) hemispheres.

Table 2.1 shows the p-values obtained from the unpaired t-test after each comparison between each pair of conditions over electrode C3. They are marked as < 0.01 (very significant); < 0.05 (significant) or the true value for > 0.05. In table D.1 the corresponding p-values for the same comparisons over electrode C4 are presented. It can be observed that, for almost all tests and subjects, except for subject 2, which exhibits no significance difference in most of the cases, that simple motor imageries in their contralateral hemisphere are comparable with the combined task in the corresponding side; and that simple motor imageries in their ipsilateral hemisphere are comparable to the rest condition in the corresponding side.

Table 2.1: Unpaired T-test to evaluate significance level at 5% of the different classes at electrode C3

Subject	RST-RH	RST-LH	RST-BHs	RH-LH	RH-BHs	LH-BHs
1	$p < 0.05$	$p < 0.05$	$p < 0.01$	$p = 0.75$	$p = 0.27$	$p < 0.13$
2	$p = 0.20$	$p = 0.25$	$p = 0.81$	$p < 0.01$	$p = 0.15$	$p = 0.40$
3	$p < 0.01$	$p = 0.80$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$
4	$p < 0.01$	$p = 0.80$	$p < 0.01$	$p < 0.01$	$p = 0.80$	$p < 0.01$
5	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p = 0.26$	$p = 0.56$	$p < 0.05$

Table 2.2: Unpaired T-test to evaluate significance level at 5% of the different classes at electrode C4

Subject	RST-RH	RST-LH	RST-BHs	RH-LH	RH-BHs	LH-BHs
1	p = 0.17	p < 0.01	p < 0.01	p < 0.05	p < 0.01	p = 0.37
2	p = 0.48	p = 0.33	p = 0.55	p = 0.07	p = 0.15	p = 0.65
3	p = 0.20	p = 0.12	p < 0.01	p < 0.01	p < 0.01	p < 0.01
4	p = 0.39	p < 0.01	p < 0.01	p < 0.01	p < 0.01	p = 0.26
5	p = 0.10	p < 0.01	p < 0.01	p < 0.01	p < 0.05	p = 0.71

Figure 2.15 shows a box plot comparing the power spectrum for subject 4 at 12 Hz of each motor imagery over electrodes C3 (fig. 2.15-A) and C4 (fig. 2.15-B). It can be noticed that there is no significant difference between the rest condition and the simple motor imageries at their ipsilateral hemisphere, nor is there for the combined task and simple motor imageries at their contralateral hemisphere.
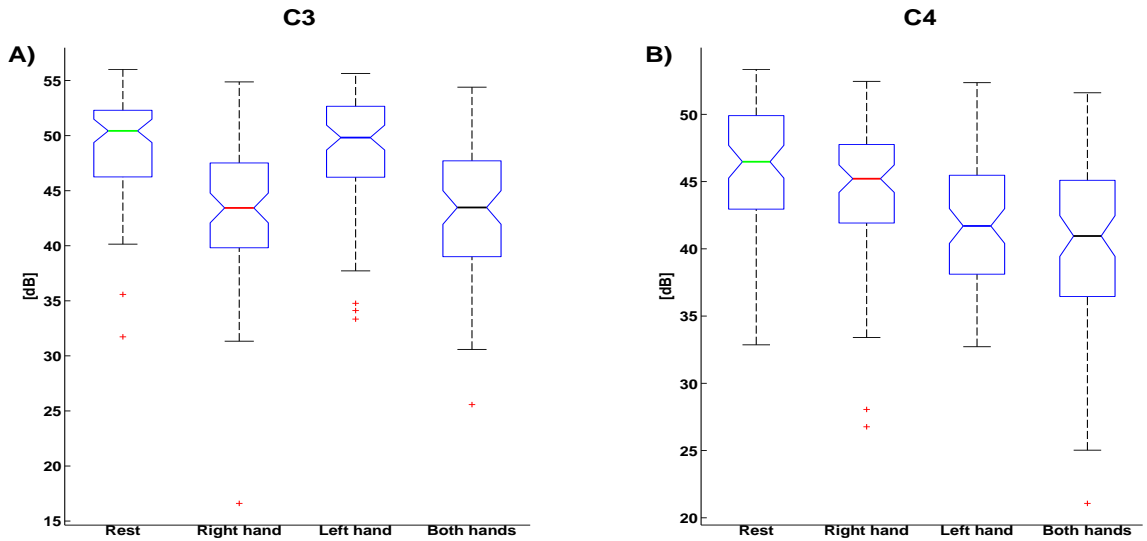
Fig. 2.15: Box plots for subject 4 comparing the power spectrum magnitude at 12 Hz among all the considered conditions over electrodes C3 (A) and C4 (B). EEG signals were taken within the time interval from [1-2 s]. The distribution of data shows that there is no significant difference between simple hand MIs at their contralateral hemisphere with the combined task in the corresponding side, nor is there between simple hand MIs in the ipsilateral hemisphere with the rest condition.

The statistical results presented in tables 2.1 and D.1, together with figure 2.15, are consistent with what has been already observed along section 1.6.2; where we have showed, in a qualitative way, the topographic similarities of the combined motor imageries involving both hands with each type of simple hand motor imageries at their contralateral side. In this way, the statistical analysis can be used to establish that EEG data recorded during a combined motor imagery

can be characterized as the superposition of the activity that is independently generated by the sources associated to the body parts engaged in the task, which confirms the hypothesis stated at the beginning of this chapter, and upon which we will develop the multilabel classification models contributed by the present work.

3

Robotic arm control

Contents

As will be pointed out below, the four different brain states that are generated by a paradigm such as the one described in the previous chapter (see section 2.6.1), are insufficient to afford an appropriate control for this kind of applications. Because of this reason, throughout the first section of this chapter we will present the 8-*class database*; which comprises EEG signals from 11 healthy subjects inducing the eight mental states that it is possible to generate by the motor tasks involving the use of the right hand, left hand, both feet together, and rest. Similarly as with the 4-*class* database, we will present a qualitative analysis on the ERD patterns that are found in the EEG signals during the eight different tasks. Furthermore, a statistical comparison of the power spectrum observed at each electrode among each pair of conditions will be presented in order to justify the grouping of classes, upon which the feature extraction methods contributed by the present work were implemented.

3.1 8-class database

EEG signals were registered at 256 Hz from 11 right-handed healthy subjects (7 men and 4 women, from 22 to 45 years old) using a commercial amplifier Refa developed by TMS International™.

Both the signal acquisition and the stimulation process were implemented on the OpenViBE platform[1] [85]. The EEG cap was fitted with 26 electrodes, namely, Fp1; Fp$_z$; Fp2; F$_z$; FC5; FC3; FC1; FC$_z$; FC2; FC4; FC6; C5; C3; C1; C$_z$; C2; C4; C6; CP5; CP3;CP1; CP$_z$; CP2; CP4; CP6, and P$_z$, re-referenced with respect to the common average reference across all channels and located over the extended international 10-20-system positions to cover the primary sensorimotor cortex (see Figure 2.7).

All experiments were carried out with the informed consent agreement of each participant and following the statements of the WMA declaration of Helsinki on ethical principles for medical research involving human subjects.

3.1.1 Paradigm and time scheme

Subjects were seated in a comfortable chair with the arms at their sides in front of a screen showing the task cue to be performed, which consisted of one of the eight mental states that it is possible to generate with all combinations including the use of the right hand, left hand, both feet together, and rest (i.e., right hand; left hand; feet; both hands together; left hand in combination with feet; right hand in combination with feet; both hands together with feet, and rest). The whole session consisted of 3 runs, containing each one 10 trials per task, which totals 30 trials per class (240 trials considering the 8 classes).

For the stimulation, three panels were simultaneously displayed on the screen, which were associated from left to right, to the left hand, feet and right hand. Each trial was randomly presented and lasted for 12 seconds, starting at second 0 with a cross at the center of each panel and an overlaid arrow indicating for the next 6 seconds the task to be performed: an arrow pointing to the left side on the left panel for left hand, an arrow pointing down on the central panel for feet, an arrow pointing to the right side on the right panel for right hand, and the simultaneous combination of all of them for the corresponding combined tasks (see Figure 3.1 and column 2 of Figure D.2). The rest condition was also considered and it was indicated by the absence of arrows. After second 6, the task cue disappeared and the crosses were remaining for the next 6 seconds indicating the pause period before the next trial started.

3.1.2 Oscillatory power analysis

Similarly to the analysis presented along section 2.6.2 in page 27, trials were considered 2 seconds before the task cue appeared on the screen and 6 seconds after it disappeared, so that each segment corresponds to a 14s-length signal segment over 26 electrodes. All trials were filtered using a 5th-order Butterworth filter within the mu frequency band, in conformance to the nature of SMRs described in the same section.

The time course of ERD/ERS% of each segment was computed by squaring its samples and smoothing the resulting signal through a 2-second-sliding window with a 125 ms shifting step

---

1. `http://openvibe.inria.fr`
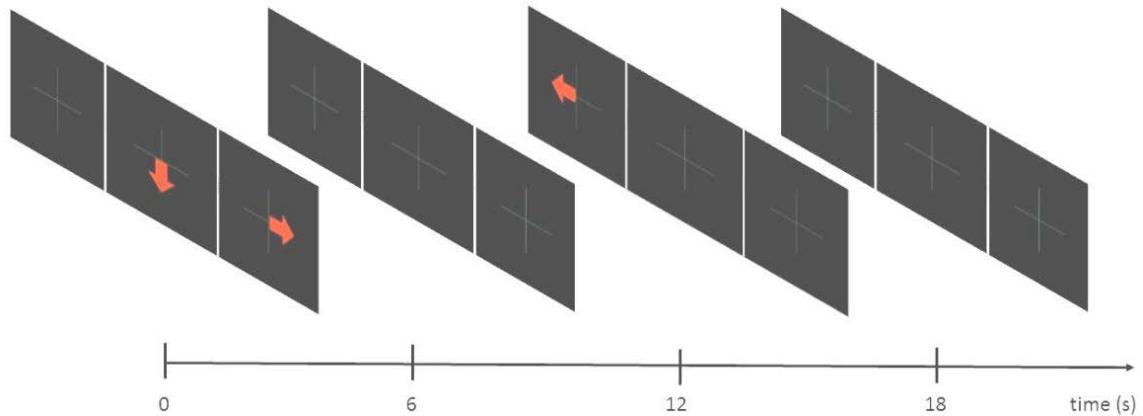
Fig. 3.1: Time scheme. Each trial was randomly presented and lasted for 12 seconds. During the first 6 seconds, users were asked to perform the motor imagery indicated by the task cue, which was launched throughout three simultaneously displayed panels associated, from left to right, to left hand, feet and right hand. The use of each body part was indicated by the presence of arrows: an arrow pointing to the left side on the left panel for left hand, an arrow pointing down on the central panel for feet, an arrow pointing to the right side on the right panel for right hand, and the simultaneous combination of all of them for the corresponding compound motor imageries. After second 6, the task cue disappeared and the crosses were remaining for the next 6 seconds indicating the pause period before the next trial started.

according to equation 2.1 (see Figure 2.6). Figure 3.2 shows a topographic representation for one of the subjects along all conditions of the ERD/ERS% values obtained for all electrodes and averaged over the 4-second period starting one second after the task cue was presented. It can be observed that the lowest values appear on the regions associated to the limbs that are engaged at each motor task (broadly, right hand over electrode C3, feet over electrode C_z, and left hand over electrode C4). The ERD/ERS% topographic maps for all subjects are presented in appendix B.
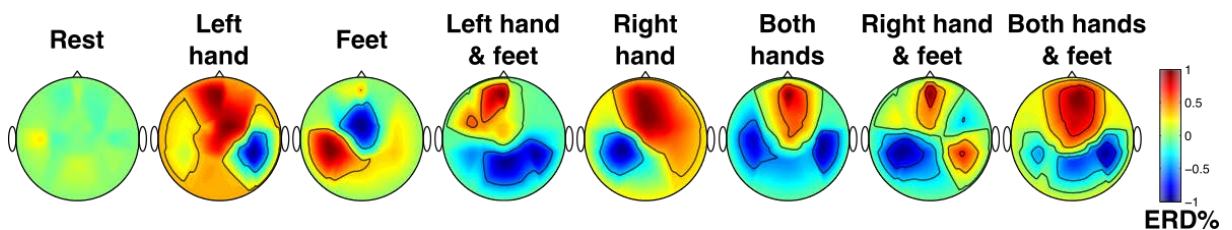
Fig. 3.2: Distribution of the relative oscillatory power along the sensorimotor cortex during the course of the different motor tasks for subject 3. ERD/ERS% values were obtained for all electrodes within the mu frequency range and averaged over the 4-second period starting one second after the task cue was presented. Negative values indicate ERD% modulation, which is characteristic of a motor task execution that appears over the contralateral hemisphere of the body part used in the process (broadly, electrode C4 for left hand, C_z for feet, and C3 for right hand). Note that each topographic map is independently normalized to enhance the corresponding patterns.

If we take a closer look over the time course of ERD/ERS% over electrode C3 (i.e, the activity source associated with the use of the right hand) presented in Figure 3.3, we can observe that all motor imageries involving the use of the right hand (i.e., right hand, right hand in combination with feet, both hands, and both hands in combination with feet) present considerably lower ERD/ERS% values than those that do not include it (i.e., rest, left hand, feet, and left hand in combination with feet).

Similarly, Figure 3.4 presents the time course of ERD/ERS% over electrode C_z (i.e, the activity source associated with the use of the feet). Over this region it can also be observed that the motor imageries involving the use of the feet (i.e., feet, left hand in combination with feet, right hand in combination with feet, and both hands in combination with feet) present lower ERD/ERS% values than those that do no it include it (i.e., rest, left hand, right hand, and both hands). However, in comparison to the contrast observed over electrode C3 between the tasks involving the use of the right hand and those that do not involve it, differences are not as significant. In fact, we can observe that the use of both left and right hands induce some desynchronization over the central region, possible because the associated sources are quite close to electrode C_z, which might cause it to incorporate some of this activity.

Finally, the same comparison is illustrated in Figure 3.5 for the time course of ERD/ERS% generated over electrode C4 (i.e, the activity source linked to the use of the left hand). In this case, similarly to the analysis over electrode C3, it is possible to observe a significant contrast between the low ERD/ERS% values obtained from the activity induced by the motor imageries involving the use of the left hand (i.e., left hand, left hand in combination with feet, both hands, and both hands in combination with feet) from the high ERD/ERS% values associated to the tasks that do not involve the use of the left hand (i.e., rest, feet, right hand, and right hand in combination with feet).

3.1.3   Statistical analysis

From the previous analysis and in correspondence to what we have reported in section 2.6.2 for the case of the 4-*class* database, it can be observed that the oscillatory power around the main activity sources associated to the body parts included in this paradigm presents similar behaviour when the corresponding limbs are engaged in a task, independently from the fact of whether the motor imagery is simple or combined. In order to formalize this analysis in a quantitative manner, we have performed the same procedure described in section 2.6.3 for evaluating the correlation between electrodes for all comparisons among each pair of motor imageries. To this end, all trials were band-pass filtered by using a 5th order Butterworth filter within the frequency range from 8 to 25 Hz (e.g., mu-band + beta-band, since ERD% values appear over both ranges). Subsequently, all segments were selected within the time interval from [2-5 s], given that for all subjects the lowest ERD% values were found over this period, within which the power spectrum was computed across all channels. Finally, the mu-beta range was scanned by taking subintervals of 2 Hz each to compute, in terms of R^2, the correlation between the power spectrum magnitude over all
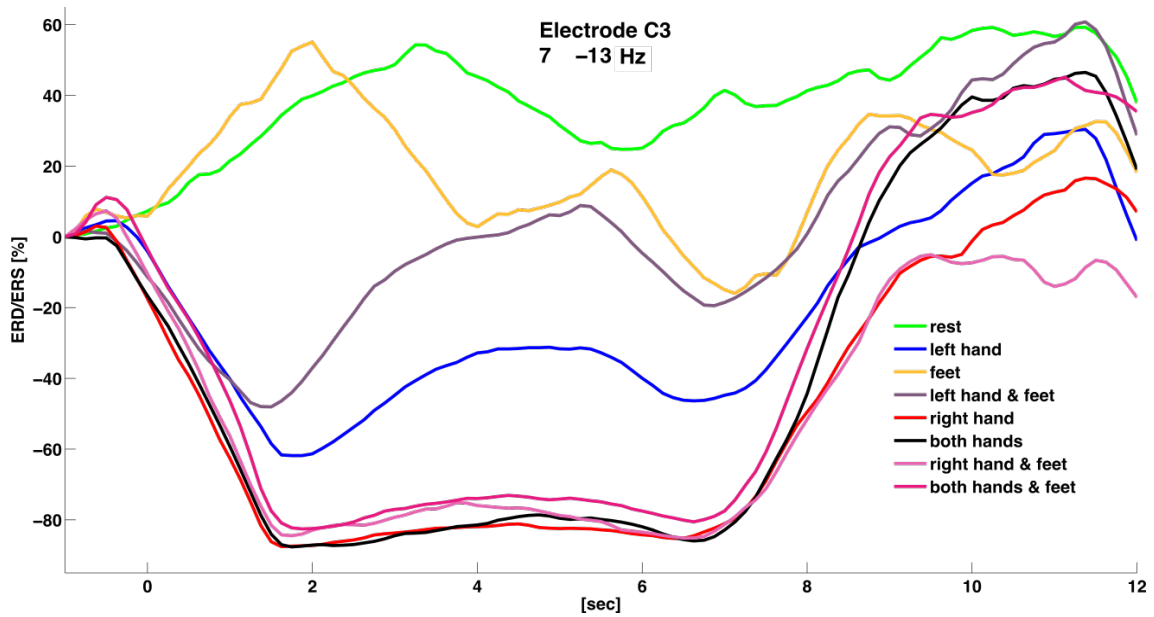
Fig. 3.3: ERD/ERS% time series within the mu range for all motor tasks over electrode C3 (i.e, the activity source associated with the use of the right hand). ERD/ERS values were estimated according to equation 2.1 over the EEG trials from subject 3. Note that all motor imageries involving the use of the right hand (i.e., right hand, right hand in combination with feet, both hands, and both hands in combination with feet) present considerably lower values that those that do not include it (i.e., rest, left hand, feet, and left hand in combination with feet).

pairs of conditions at each electrode [4]. Note that for the analysis over the trials contained in the 8-*class* database we have selected a wider time window than the one that was used for the 4-*class* database. This is due to the morphology that was observed for this set of subjects, which in general presented longer desynchronization periods.

Similarly to what we have found in the case of the 4-*class* database, for all subjects the largest R^2 values were located around 12 Hz. Figure 3.6 shows the topographic maps generated for one of the subjects throughout all comparisons between the power spectrum magnitudes at this frequency for each pair of conditions.

High R^2 values indicate that there is a significant difference between the considered motor tasks over the corresponding positions. On the contrary, small R^2 values show that the conditions under analysis do not present a significant difference, and therefore are similar between each other [21]. It can be observed that the highest R^2 values occur when the motor imageries differ on the limbs that are engaged during their execution. For instance, if we analyze the first row of Figure 3.6, which corresponds to the comparison between the power spectrum calculated during rest and each one of the values generated by the other motor tasks, we notice that the most significant differences occur just around the activity sources associated to the limbs that are engaged during each one of the motor imageries. This occurs because during rest none of the sources induces ERD modulation, whereas for the other tasks this activity appears over the associated sources, which causes such differentiation. The same behaviour follows for all pair of comparisons, which
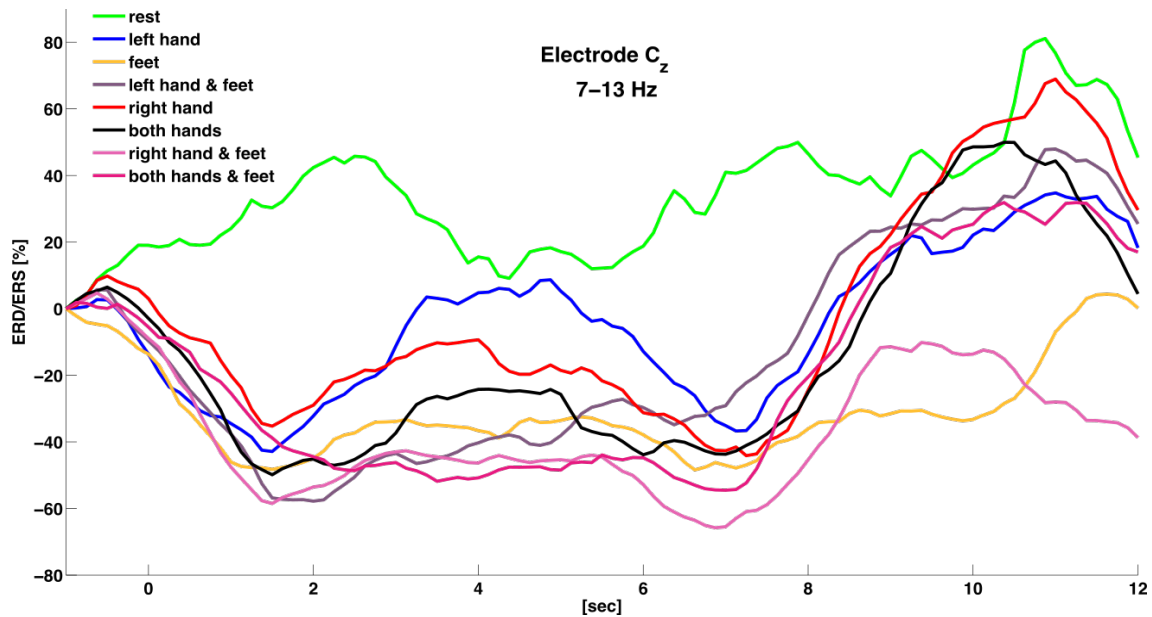
Fig. 3.4: ERD/ERS% time series within the mu range for all motor tasks over electrode C_z. ERD/ERS values were estimated according to equation 2.1 over the EEG trials from subject 3. Note that the motor imageries involving the use of the feet (i.e., feet, left hand in combination with feet, right hand in combination with feet, and both hands in combination with feet) present lower ERD/ERS% values than those that do no it include it (i.e., rest, left hand, right hand, and both hands). However, this contrast is not as significant as the one observed on electrode C3, possibly due to the fact that the regions associated to both the left and right hands are quite close to electrode C_z, which might cause it to incorporate some of this activity.

coincides with what it was observed over the ERD/ERS% time series from the previous section.

Additionally to the correlation analysis between conditions, we present a series of box plots for comparing the power spectrum magnitudes at 12 Hz between all motor tasks over the regions associated to the main sources. Again, it is possible to observe that motor imageries involving the use of the limb associated to each electrode present similar low power values at these positions, whether those that are not engaged generate higher values. This is particularly visible from figure 3.7, which presents the corresponding analysis over electrode C3 (i.e., the region associated to the right hand). Note that the distribution of the mean values confirm that there is no significant difference between motor imageries including the use of the right hand over this region, independently from the fact of whether the motor imagery is simple or combined. In contrast there is a significant difference with all the other tasks that do not include the use of the right hand which, in turn, present a uniform distribution between them. The same analysis is presented for electrode C_z (i.e., the region associated to the feet) and electrode C4 (i.e., the region associated to the left hand) in figures 3.8 and 3.9 respectively.
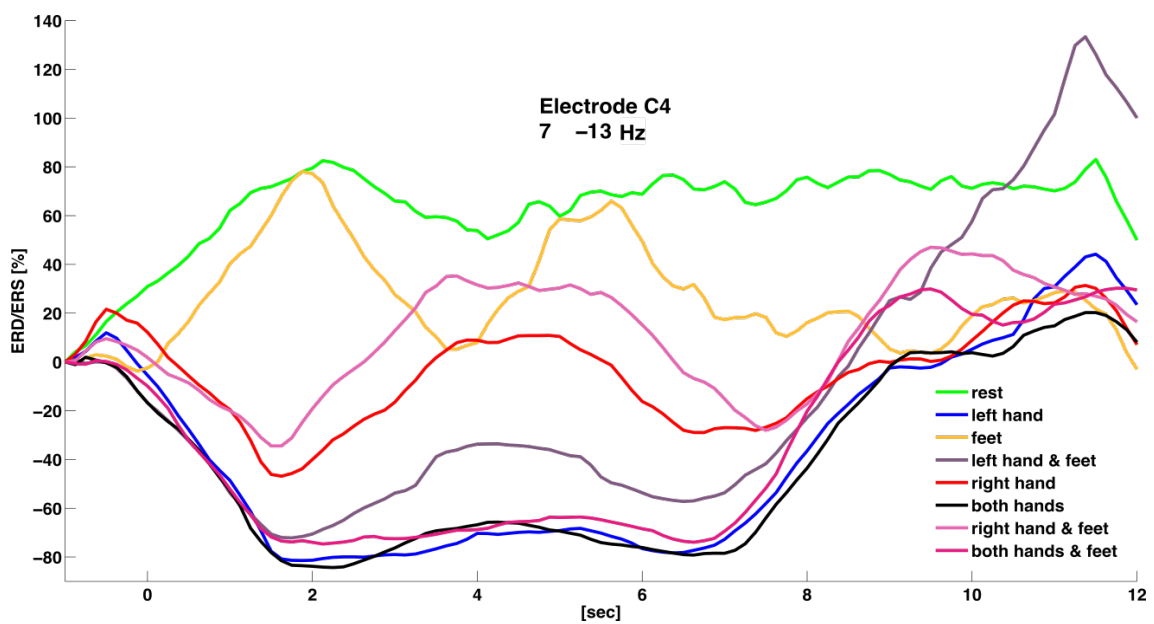
Fig. 3.5: ERD/ERS% time series within the mu range for all motor tasks over electrode C4. ERD/ERS values were estimated according to equation 2.1 over the EEG trials from subject 3. Note that all motor imageries involving the use of the left hand (i.e., left hand, left hand in combination with feet, both hands, and both hands in combination with feet) present considerably lower values that those that do not include it (i.e., rest, right hand, feet, and right hand in combination with feet).
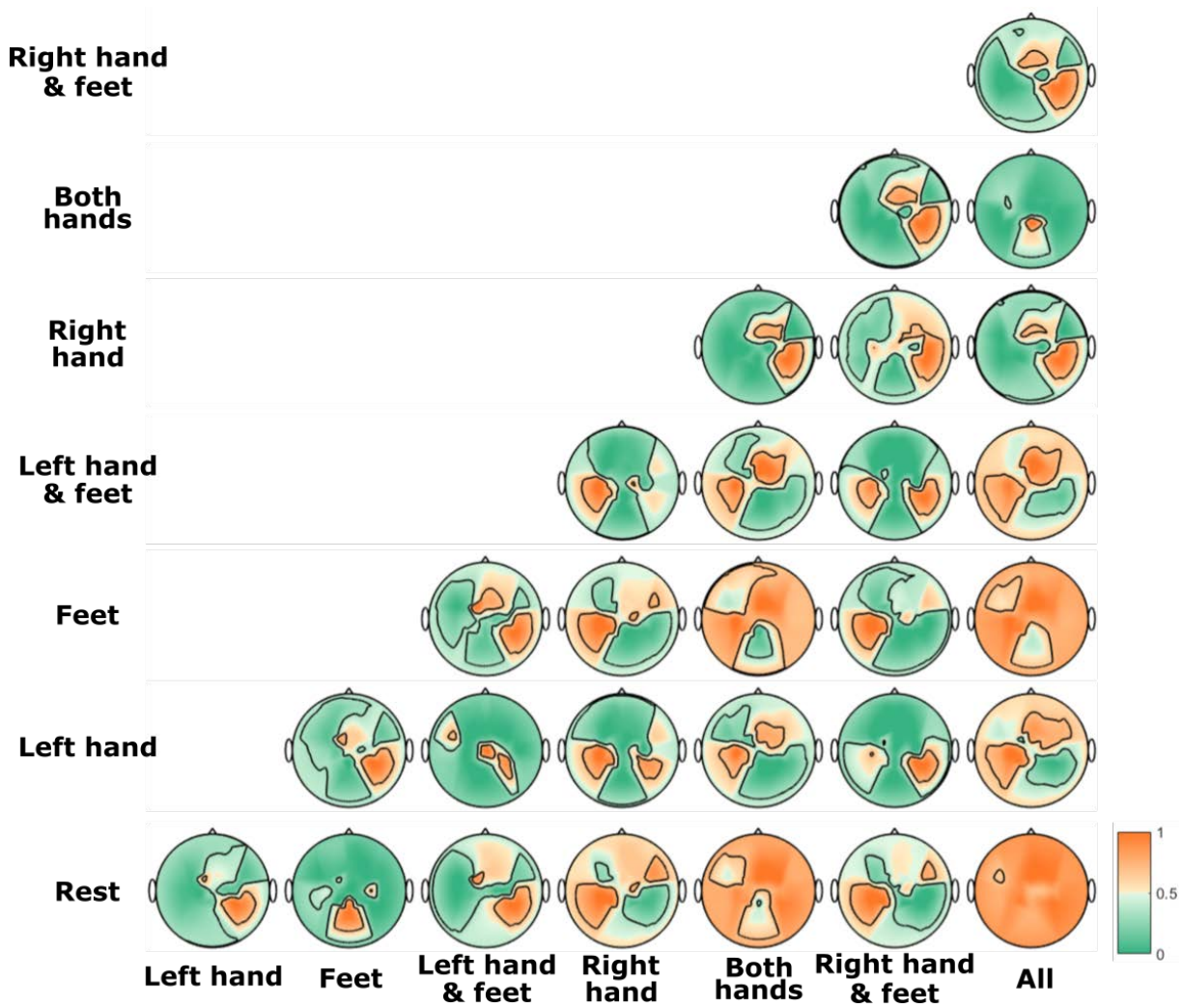
Fig. 3.6: Topographic representations for one of the subjects of the correlation values between the comparison over all conditions in terms of R^2. High R^2 values indicate that there is a significant difference between the considered motor tasks over the corresponding positions. Note that each topographic map is independently normalized to enhance the corresponding patterns.
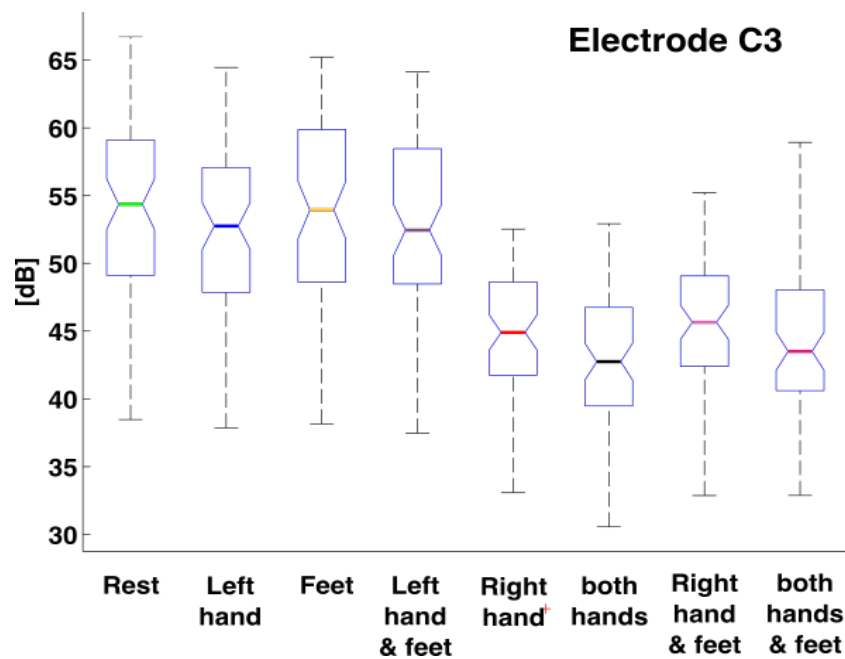
Fig. 3.7: Box plots for subject 3 comparing the power spectrum magnitude at 12 Hz within the time interval from [2-5 s] among all the considered conditions over electrode C3. The distribution of data shows that there is no significant difference between motor imageries including the use of the right hand over this region, independently from the fact of whether the motor imagery is simple or combined. In contrast there is a significant difference with all the other tasks that do not include the use of the right hand which, in turn, present a uniform distribution between them.
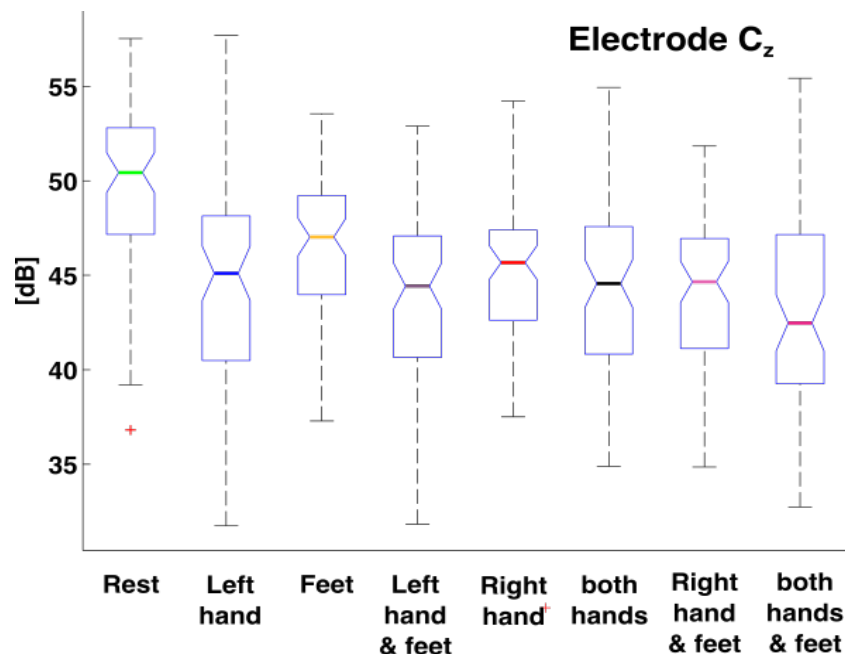
Fig. 3.8: Box plots for subject 3 comparing the power spectrum magnitude at 12 Hz within the time interval from [2-5 s] among all the considered conditions over electrode C_z. The distribution of data reflects what was observed from the ERD/ERS% time series, and confirms that EEG activity around the central region is influenced by the surrounding sources.
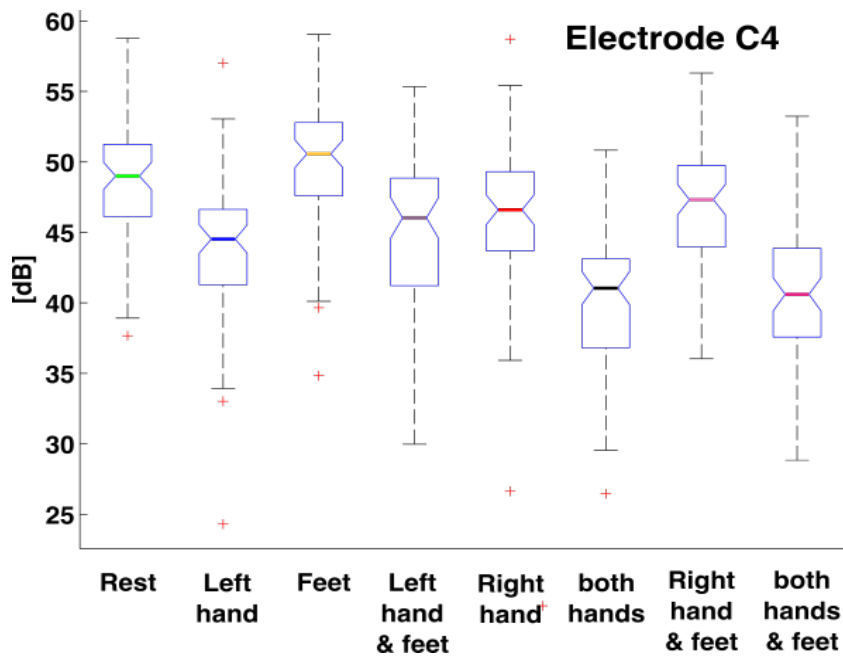
Fig. 3.9: Box plots for subject 3 comparing the power spectrum magnitude at 12 Hz within the time interval from [2-5 s] among all the considered conditions over electrode C4. The distribution of data shows that there is no significant difference between motor imageries including the use of the left hand over this region, independently from the fact of whether the motor imagery is simple or combined. In contrast there is a significant difference with all the other tasks that do not include the use of the left hand which, in turn, present a uniform distribution between them.

4

Signal Processing

Contents

As previously mentioned, a BCI system aims at recognizing determined patterns in the user's brain activity that are related to a mental task that the person performs in order to transmit the associated command. To this end two central stages are required, namely, feature extraction and feature classification. The first concept consists in determining the characteristics that confer distinctiveness to the brain activity that is elicited during the corresponding task, so that it can

be recognized from other states. In most of the cases this process reduces the dimensionality of data by excluding redundant information, and it enhances the attributes of the signal by exploiting the aspects that are related to the physiological phenomena. The derived values are projected into the feature space, where elements from different conditions are expected to be located as separated as possible, so that distinct groups can be easily identified. Once features have been extracted from the original input data, the next step consists in creating a classification algorithm (classifier) that is able to categorize such elements according to their position within the classification space, so that the system can proceed to translate the corresponding command.

In the following chapter the concept of feature extraction is introduced and elaborated in the context of EEG signals during motor imagery. To this end, a widely used algorithm known as Common Spatial Pattern (CSP), which has been shown to be powerful at discriminating sensorimotor rhythms and lays the basis for the methods contributed by the present work, is elucidated together with two of its variants, namely, Analytic Common Spatial patterns (ACSP) and Common Spatial Pattern by Joint Approximate Diagonalization (CSP by JAD). Additionally, we will present a strategy called Filter Bank Common Spatial Pattern (FBCSP) that optimizes the algorithm performance by decomposing the EEG signals into a series of frequency bands, which favors the selection of a discriminative subject-specific frequency range. In this way, the dimensionality of the feature space is drastically increased and some redundant information is generated. At this point we will introduce the concept of *feature selection*, which allows choosing the most descriptive elements, so that computational costs are optimized and models are prevented of considering information that might interfere with discriminative patterns. In this sense we will present two mutual-information-based methods for feature selection, namely, the Mutual Information-based Best Individual Feature (MIBIF) approach, and the Information Theoretic Feature Extraction (ITFE) algorithm. The first one is applied together with the FBCSP method, which represents a powerful solution that has drawn the attention of the community after winning the BCI competition IV [86]. The second one is presented in combination with the CSP by JAD algorithm as an non-heuristic strategy to select the filters that result from the joint diagonalizaton of multiple covariance matrices.

Since the principle of the CSP algorithm is formulated upon the information contained within the covariance matrices of the EEG signals, we will first provide an insight into their definition and interpretation. Subsequently, along the same path, we will formulate the eigenvalue problem that solve the simultaneous diagonalization of the covariance matrices related to two conditions, which represents the foundation of the CSP algorithm and provides the contrast that is useful for the further classification.

The second section of this chapter is dedicated to present the notion of classification, for which two commonly used algorithms will be described, namely, Linear Discriminant Analysis (LDA) and Support Vector Machines (SVM). These classification algorithms are based on a discriminant function, which implies that they formulate a relation that maps each feature vector directly onto a class label. There are other classification methods that do not use discriminant

functions to categorize input elements, such as neural networks, decision trees and, among others, distance-based algorithms. The latter approach has drawn special attention due to the possibility of applying it in the context of Riemannian geometry, which has also demonstrated to be potent at differentiating sensorimotor rhythms. In this regard and, in order to explore other options, the present work considers as well this solution and provides a brief outline as an alternative to the CSP algorithm; which endows a classification that can be directly achieved within the manifold of symmetric and positive definite matrices given by the covariance matrices of the EEG signals that are considered as descriptors, from where it is possible to extract the spatial information without using spatial filtering, and thus simplifying the classification process.

Finally, the last section of the chapter introduces a new classification method for motor imagery-based BCI systems that makes use of the combination of the CSP algorithm and Riemannian geometry, which represents one of the contributions of the present work. The motivation of this approach lies in the fact that the covariance matrices of the projected signals preserve the differences attained through the spatial filtering, making it advantageous to apply the CSP algorithm before mapping them into the Riemannian space, so that more discriminative features are generated.

4.1 Feature extraction

A fundamental machine learning task consists in searching for patterns in data; which can be achieved by tuning the parameters of an adaptive model that incorporates the characteristics of a data set that is called *training* set. The categories (or classes) of the elements contained in the training set are known in advance, typically as the result of conducting a supervised experiment. These categories can be expressed by a target vector $\mathbf{t}$ representing the identity of each element $\mathbf{x}$. Then, the aforementioned model can be expressed as a function $\mathbf{y(x)}$; which takes a new element $\mathbf{x}$ as input to generate an output vector $\mathbf{y}$ that is encoded in the same manner as the target vectors. The specific form of the function $\mathbf{y(x)}$ is determined by the training set during the learning phase. Once the model is trained, it is possible to determine the identity of new elements that are said to compose the *testing* set, which must be processed in the same way as the training data. In most of the cases, the variability of the input vectors will be such that the training set can not represent the entire population, and thus the task to correctly categorize new elements differing from those used in the training phase becomes a challenge.

For many applications, the original input vectors are transformed into a new space of variables where, hopefully, the pattern recognition problem will be easier to solve. This processing stage is called feature extraction and, basically, it consists of creating derived values (features) from the original input vectors that are meant to be informative and non-redundant, so that the subsequent generalization and interpretation are facilitated due to the reduced representation of the complete initial training set. It is worth mentioning that not all methods for feature extraction reduce dimensionality, some of them aim at improving the signal-to-noise ratio without affecting data

dimension, whereas others still increase it towards a better class separability [87].

4.1.1 Covariance matrix

Let $\mathbf{x} \in \mathbb{R}^n$ be a random vector with multiple dimensions containing a scalar random variable within each element, then its covariance matrix is defined by the matrix whose element in the (i, j) position corresponds to the covariance between the ith and j^{th} elements of $\mathbf{x}$. In an intuitive manner, the covariance matrix generalizes the concept of covariance to multiple dimensions. For instance, if the elements of the column vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

are random variables, each with finite variance, then the covariance matrix $\boldsymbol{\Sigma}$ is the matrix whose (i,j) entry is given by

$$\boldsymbol{\Sigma}_{i,j} = \mathrm{cov}(x_i, x_j) = \mathrm{E}[(x_i - \mu_i)(x_j - \mu_j)^{\top}]$$

where

$$\mu_i = \mathrm{E}[x_i]$$

is the expected value of the ith element of vector $\mathbf{x}$. In other words,

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathrm{E}[(x_1 - \mu_1)(x_1 - \mu_1)] & \mathrm{E}[(x_1 - \mu_1)(x_2 - \mu_2)] & \ldots & \mathrm{E}[(x_1 - \mu_1)(x_n - \mu_n)] \\ \mathrm{E}[(x_2 - \mu_2)(x_1 - \mu_1)] & \mathrm{E}[(x_2 - \mu_2)(x_2 - \mu_2)] & \ldots & \mathrm{E}[(x_2 - \mu_2)(x_n - \mu_n)] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}[(x_n - \mu_n)(x_1 - \mu_1)] & \mathrm{E}[(x_n - \mu_n)(x_2 - \mu_2)] & \ldots & \mathrm{E}[(x_n - \mu_n)(x_n - \mu_n)] \end{bmatrix}$$

Since the covariance of the ith element with itself is its variance, each element on the (i,i) position within the covariance matrix diagonal corresponds to the variance of the ith element. Moreover, because $\mathrm{cov}(x_i, x_j) = \mathrm{cov}(x_j, x_i)$ every covariance matrix is symmetric. In addition, every covariance matrix is positive semi-definite, which implies that their eigenvalues are non-negative [88].

The distribution of data is reflected in their covariance matrix (see Figure 4.1). The diagonal spread is related to the covariance (off-diagonal of the matrix), whereas the dispersion in the horizontal and vertical axis is associated to the variance (diagonal of the matrix).
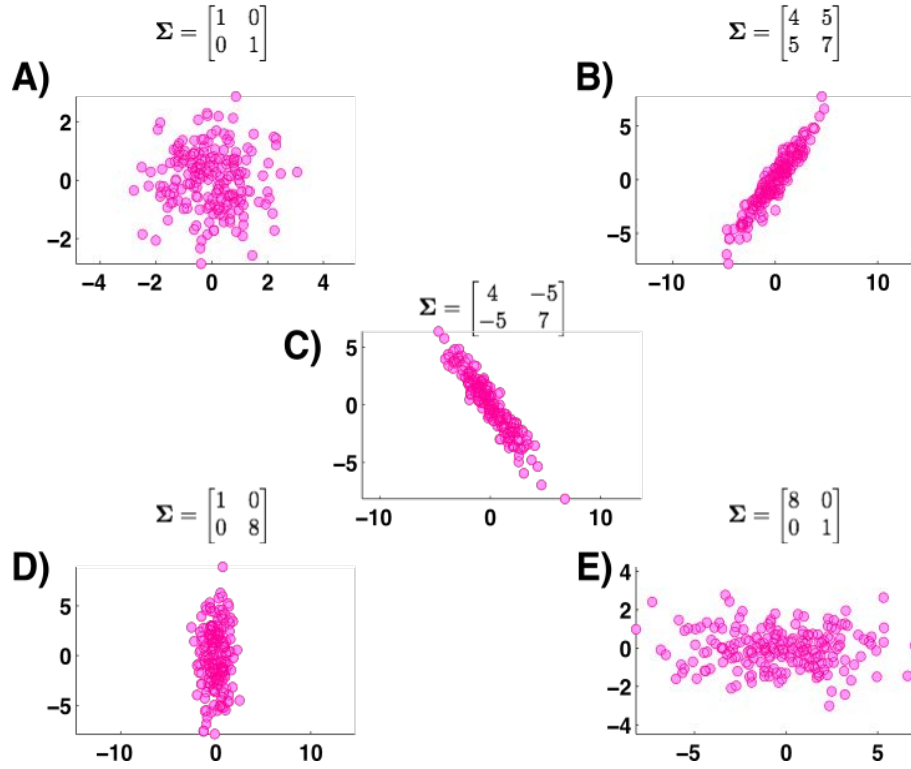
Fig. 4.1: The distribution of data is reflected in their covariance matrix. Diagonal spread is captured by the covariance, while axis-aligned spread is captured by the variance. A) shows a white noise distribution, in this case the covariance matrix is diagonal (i.e., all non-diagonal elements are zero). B) and C) present a diagonal distribution whose correlation is reflected in the sign of the non-diagonal elements of the covariance matrix. D) and E) show different dispersions in relation to the main axis, which can also be observed from their covariance matrices.

4.1.2 Normal distribution and eigenvalue decomposition

The normal distribution, also referred as Gaussian, is a widely used model for the distribution of continuous variables, which appears in several different contexts and can be motivated from many perspectives [87]. For a single variable x, the normal distribution can be written in the following terms

$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{(2\sigma^2)}(x - \mu)^2\right\}, \tag{4.1}$$

where μ and σ^2 correspond, respectively, with the mean and variance of x. In the case of a D-dimensional vector $\mathbf{x}$, the multivariate normal distribution takes the form

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\}, \tag{4.2}$$

where $\boldsymbol{\mu}$ corresponds with the D-dimensional mean vector of $\mathbf{x}$ and $\boldsymbol{\Sigma}$ with its $D \times D$ covariance matrix, whose determinant is represented by $|\boldsymbol{\Sigma}|$.

Now, the functional dependence of the normal distribution on $\mathbf{x}$, takes the quadratic form of the *Mahalanobis distance*

$$\Delta^2 = (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}), \tag{4.3}$$

which is found within the exponent of equation 4.2. The Mahalanobis distance is a measure of the distance between a point $\mathbf{x}_i$ and the distribution $\mathcal{N}$, and it represents a multidimensional generalization of the idea of measuring how many standard deviations away $\mathbf{x}_i$ is from $\boldsymbol{\mu}$ [89]. If $\mathbf{x}_i$ is at $\boldsymbol{\mu}$, then the distance is zero, and it grows as $\mathbf{x}_i$ moves away. Along each principal component axis, it measures the number of standard deviations from $\mathbf{x}_i$ to $\boldsymbol{\mu}$. If each of these axes is rescaled to have unit variance, (i.e., when $\boldsymbol{\Sigma}$ is the identity matrix), then the Mahalanobis distance reduces itself to the Euclidean distance in the transformed space.

We can consider $\boldsymbol{\Sigma}$ as a symmetric matrix without loss of generality, since any antisymmetric component would disappear from the exponent. Now consider the eigenvalue decomposition for the covariance matrix

$$\boldsymbol{\Sigma}\mathbf{u}_i = \lambda \mathbf{u}_i, \tag{4.4}$$

where $i = 1, ..., D$. Since $\boldsymbol{\Sigma}$ is a real and symmetric matrix, its eigenvalues will be real and, its eigenvectors, can be chosen to form an orthonormal set, so that

$$\mathbf{u}_i^\top \mathbf{u}_j = \boldsymbol{I}_{ij}, \tag{4.5}$$

where $\boldsymbol{I}_{ij}$ is the i, j element of the identity matrix satisfying

$$\boldsymbol{I}_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise.} \end{cases} \tag{4.6}$$

The covariance matrix $\boldsymbol{\Sigma}$ can be represented as an expansion in terms of its eigenvectors, so that

$$\boldsymbol{\Sigma} = \sum_{i=1}^{D} \lambda_i \mathbf{u}_i \mathbf{u}_i^\top, \tag{4.7}$$

then as well, the inverse covariance matrix $\boldsymbol{\Sigma}^{-1}$, can be rewritten in the form

$$\boldsymbol{\Sigma}^{-1} = \sum_{i=1}^{D} \frac{1}{\lambda_i} \mathbf{u}_i \mathbf{u}_i^\top. \tag{4.8}$$

Substituting equation 4.8 into equation 4.3, the Mahalanobis distance quadratic form becomes

$$\Delta^2 = \sum_{i=1}^{D} \frac{y_i^2}{\lambda_i}, \tag{4.9}$$

where y_i has been defined as

$$y_i = \mathbf{u}_i^\top (\mathbf{x} - \boldsymbol{\mu}) \tag{4.10}$$

The variables $\{y\_i\}$ can be interpreted as a new coordinate system defined by the orthonormal vectors $\mathbf{u}_i$ that are shifted and rotated with respect to the original x_i coordinates. By forming the vector $\mathbf{y} = (y_i, ..., y_D)^\top$ we obtain

$$\mathbf{y} = \mathbf{U}(\mathbf{x} - \boldsymbol{\mu}) \tag{4.11}$$

where $\mathbf{U}$ is a matrix whose rows are given by $\mathbf{u}_i^\top$. From equation 4.5 it follows that $\mathbf{U}$ is an orthogonal matrix (i.e., it satisfies $\mathbf{U}\mathbf{U}^\top = I$, and also $\mathbf{U}^\top\mathbf{U} = \boldsymbol{I}$), where $\boldsymbol{I}$ is the identity matrix.

The quadratic form, and hence the normal density, will be constant on surfaces for which equation 4.10 is constant. If all of the eigenvalues λ_i are positive, then these surfaces represent ellipsoids, with their centres at $\boldsymbol{\mu}$ and their axes oriented along $\mathbf{u}_i$, and with scaling factors in the directions of the axes given by $\lambda^{1/2}$, as illustrated in figure 4.2.

For the normal distribution to be well defined, it is necessary for all of the eigenvalues λ_i of the covariance matrix to be strictly positive, otherwise the distribution cannot be properly normalized. A matrix whose eigenvalues are strictly positive is said to be *positive definite*.
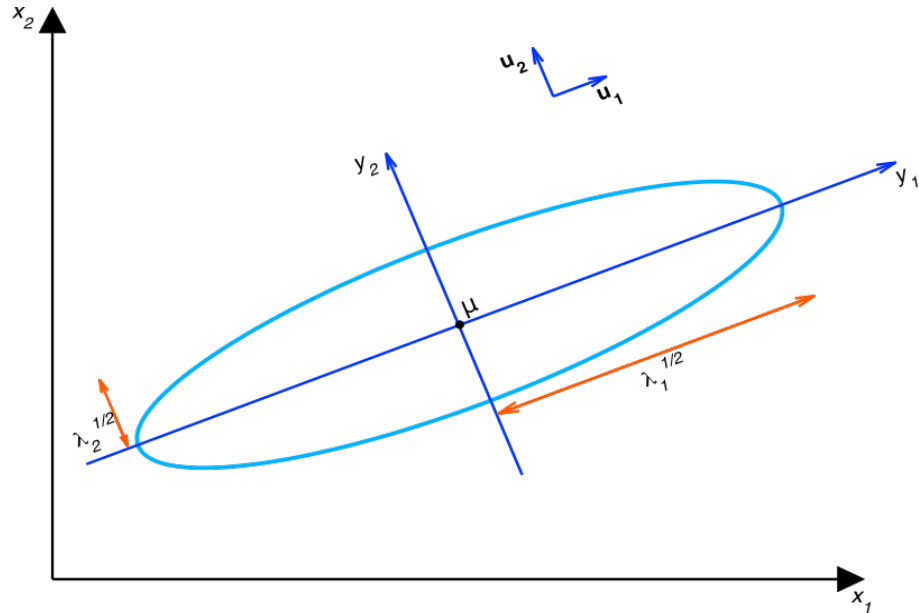
Fig. 4.2: Normal distribution. The blue curve shows the elliptical surface of constant probability density for a normal distribution within a two-dimensional space $\mathbf{x} = (x_1, x_2)$, on which the density is exp(-1/2) of its value at $\mathbf{x} = \mu$. The major axes of the ellipse are defined by the eigenvectors $\mathbf{u}_i$ of the covariance matrix, with corresponding eigenvalues λ_i.

4.1.3 Common Spatial Patterns (CSP)

A very convenient and widely used algorithm to extract motor imagery features from EEG signals is known as Common Spatial Pattern (CSP); which generates a series of spatial filters that are applied to decompose multi-dimensional data into a set of uncorrelated components [20]. These filters aim at extracting elements that simultaneously maximize the variance of one class, while minimizing the variance of the other one.

The first step of the CSP method involves the computation of the normalized covariance matrix Σ_k

$$\boldsymbol{\Sigma}_k = \frac{\mathbf{E}_k\mathbf{E}_k^\top}{\text{trace}(\mathbf{E}_k\mathbf{E}_k^\top)}, \tag{4.12}$$

where $\mathbf{E}_k \in \mathbb{R}^{C\times T}, k \in \{1,2\}$ denotes the EEG single trials belonging to class k over C channels and T samples, with $^\top$ representing the transpose. Then, the means of the covariance matrices $\overline{\boldsymbol{\Sigma}}_1$ and $\overline{\boldsymbol{\Sigma}}_2$ are computed by averaging for each class the covariance matrices of the successive training trials. So that the spatial filters $\mathbf{W}$ can be computed by solving the following generalized eigenvalue decomposition problem [20]:

$$\overline{\boldsymbol{\Sigma}}_1\mathbf{W} = \lambda\overline{\boldsymbol{\Sigma}}_2\mathbf{W} \tag{4.13}$$

with λ denoting the diagonal matrix containing the eigenvalues of $\boldsymbol{\Sigma}_1$. The spatial filtered signal $\mathbf{Y_k}$ can be obtained from the EEG trials $\mathbf{E_k}$ as

$$\mathbf{Y_k} = \mathbf{W}^\top\mathbf{E_k}. \tag{4.14}$$

Each column vector of $\mathbf{W}$ represents a spatial filter, and there are as many as the number of channels in the EEG signal. Note that the CSP filters are paired and that not all of them are relevant for the subsequent classification; therefore, after being sorted according to the values of λ (either in an ascending or descending order instinctively), only the first m and the last m columns of $\mathbf{W}$ are selected (In the present work, for all methods and subjects $m = 2$ pairs of filters were considered [90]).

The selected elements generate the spatial filters coefficient matrix $\widetilde{\mathbf{W}}$, from which the m pairs of CSP features of the ith trial for the band-pass filtered EEG measurements can be computed as

$$\mathbf{v}_i = \log\frac{\text{diag}(\widetilde{\mathbf{W}}^\top\mathbf{E}_i\mathbf{E}_i^\top\widetilde{\mathbf{W}})}{\text{trace}[\widetilde{\mathbf{W}}^\top\mathbf{E}_i\mathbf{E}_i^\top\widetilde{\mathbf{W}}]} \tag{4.15}$$

In order to illustrate the CSP algorithm, let us consider a toy example for which only the EEG signals recorded over electrodes C3 and C4 during left hand and right hand motor imageries are considered. To this end, we will use the signals from one male subject included in the 4-class database. The corresponding average across trials (n=80, see section 2.6.1), for a 4-second window starting 0.5 s after the task cue, is shown in figure 4.3-A for electrode C3, and in figure 4.3-B for

electrode C4. Note that each trial was independently filtered between the frequency range from 8 to 30 Hz by using a 5th order Butterworth filter. If we consider the EEG signals in a 2-dimensional representation, so that the samples recorded on C3 correspond to one dimension (abscissa axis) and, the samples recorded on C4 to the second dimension (ordinate axis), for the left hand and right hand motor imageries independently, we obtain the scatter plot of figure 4.3-C. From this point of view we can observe that both groups of data present a normal distribution. Moreover, the variance of each distribution exhibits the characteristics of the sensorimotor rhythms discussed in section 2.3, and presents a smaller magnitude with respect to the dimension associated to the contralateral electrode, which reflects the desynchronization of the electrical activity during the corresponding task. However, even though both distributions present ERD lateralization, data are likely correlated. In this sense, the transformation that the CSP algorithm produces by projecting the original trials over the generalized eigenvectors (see equations 4.13 and 4.14), increases the contrast between the variance of both groups, so that the smallest variance of one distribution is found along the dimension for which the other group presents the largest one, and vice versa. The projected trials after applying the CSP algorithm are shown, for electrodes C3 and C4, in figure 4.4-A and figure 4.4-B respectively. We can observe that the lateralization for each motor task is enhanced with respect to figures 4.3-A-B (consider the axes scale). Furthermore, if we look at the distribution of the transformed data (see 4.4-C), we observe that both groups are simultaneously uncorrelated.

In fact, as described at the beginning of this section, the CSP algorithm considers the whole set of electrodes that are placed during the recording session, which normally includes those covering the sensorimotor cortex. The number of resulting filters equals the number of EEG channels and, in most of the cases, at least two pairs (i.e., four filters) are selected. With this scenario the performance of the method is even better than the one observed in the previous example. Figure 4.5 shows the projections of two consecutive single trials for right hand and left hand motor imageries over the first two and the last two CSP filters (from up to down) computed with equation 4.14. In order to obtain the spatial filters coefficient matrix, all trials were considered over the whole set including the 26 electrodes used in the setup (see figure 2.7 in section 2.6.1), and by using a 4-second window starting 0.5 s after the task cue. All trials were previously and independently filtered between the frequency range from 8 to 30 Hz by using a 5th order Butterworth filter. We can observe that the signals variance is significantly increased during the task period for one type of motor imagery while it is decreased for the other one when using one pair of CSP filters and, the same behavior, but in the other way around, when using the other pair.

Because of the manner in which the CSP algorithm achieves such a contrast by the simultaneous increasing/decreasing of the signals variance, this approach is only suitable to discriminate between two classes and, in order to solve problems involving many classes, multiclass extensions are required [2], [15]. Along this chapter we will introduce a multiclass approach based on the joint approximate diagonalization (JAD) of the mean covariance matrices of multiple classes involved
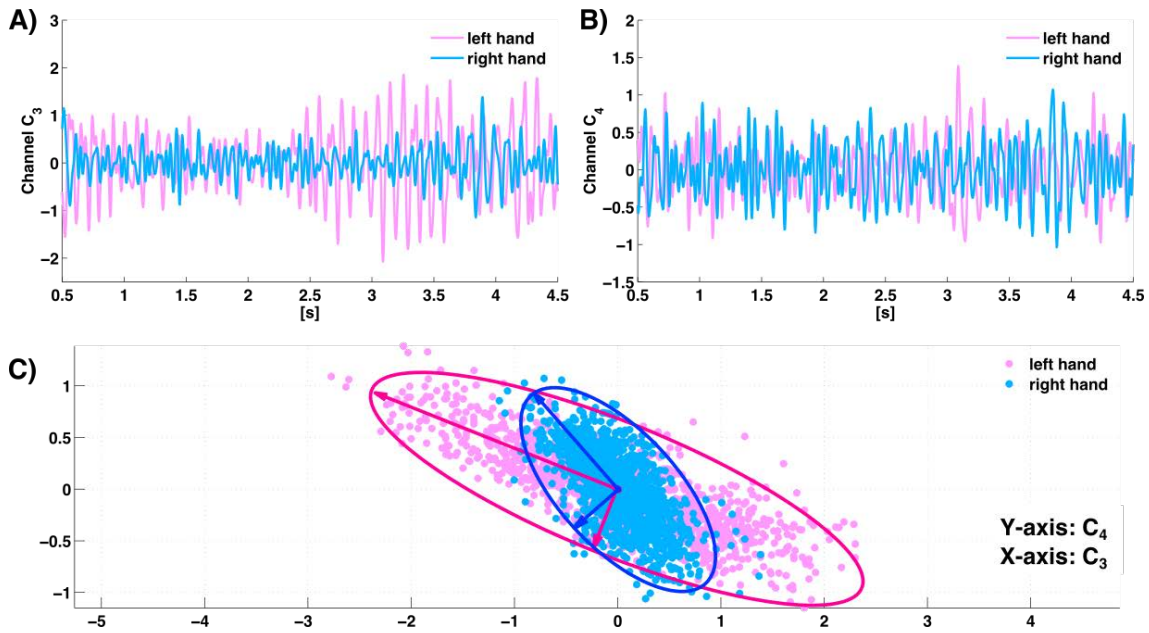
Fig. 4.3: EEG trials during left hand and right hand motor imageries in electrodes C3 (A) and C4 (B). Figure (C) shows a 2D representation where it is possible to observe that both groups are normally distributed. The variance of each distribution exhibits the characteristics of the sensorimotor rhythms, and presents a smaller magnitude with respect to the dimension associated to the contralateral electrode, which reflects the desynchronization of the electrical activity during the corresponding task. However, even though both distributions present ERD lateralization, data are likely correlated.
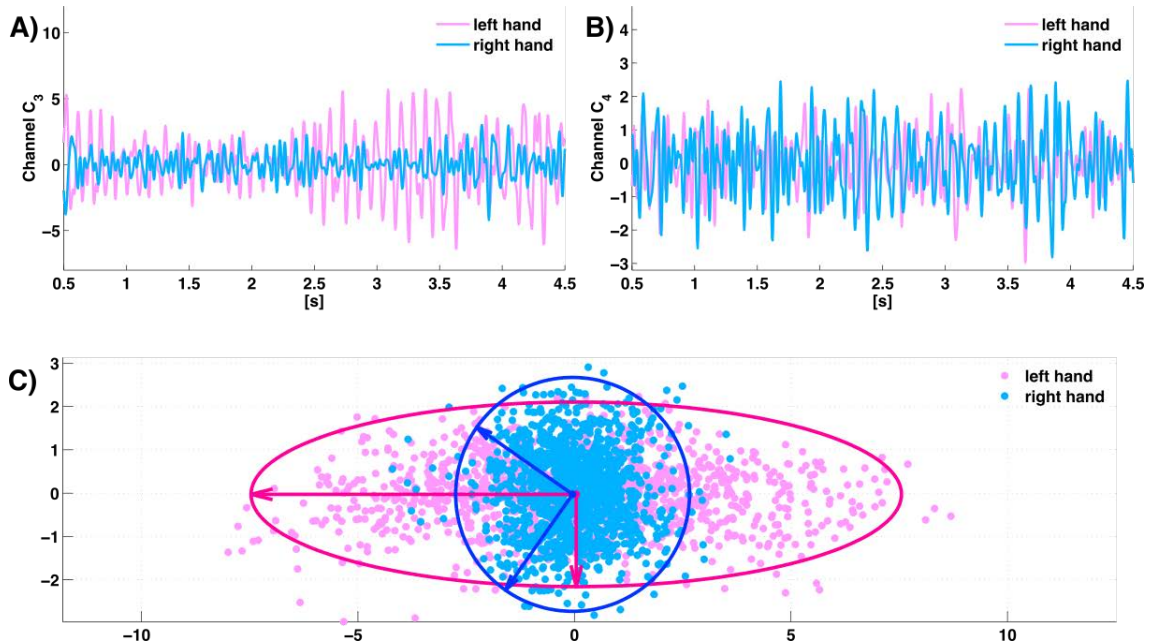
Fig. 4.4: EEG trials during left hand and right hand motor imageries in electrodes C3 and C4 after having applied the CSP algorithm, which generates uncorrelated data for both classes.
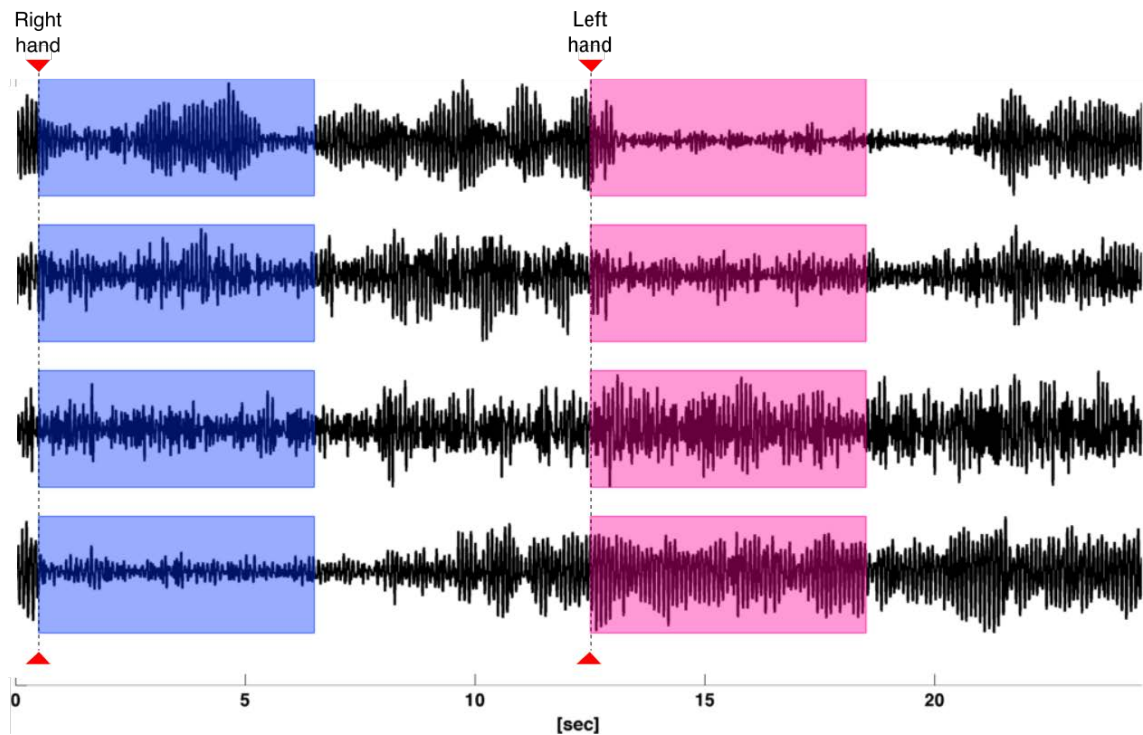
Fig. 4.5: Projected trials generated by the first two CSP pairs of spatial filters over the band-pass filtered EEG signals of two consecutive right hand and left hand motor imageries. From top to down, the first and last projections are associated to the first CSP filter, while the two in the middle are linked to the second pair. Note the simultaneous changes of the signals variance within each pair of projections.
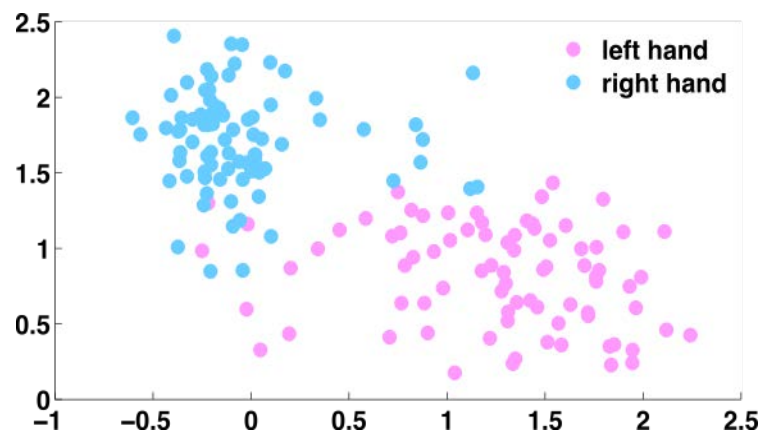
Fig. 4.6: CSP features extracted from the changes in variance of the projected trials according to equation 4.15

in a classification task, which is the same principle of the eigenvalue decomposition discussed here (see section 4.5).

4.2 Analytical Common Spatial Patterns (ACSP)

In contrast to the standard CSP algorithm, ACSP considers data in its complex form, which can provide more information that when considering only the real-value counterpart [91]. Without this, both methods aim at discriminating between two classes of data by finding a set of spatial filters that simultaneously maximize the variance of one class, while minimizing the variance of the other one. The ACSP method starts by converting the band-passed EEG signals $s(t)$ into their complex-valued counterparts $\widetilde{s}(t)$, which can be achieved through the computation of the Hilbert transform in the following terms

$$\widetilde{s}(t) = \frac{1}{\pi} p.v. \int \frac{s(\tau)}{t - \tau} d\tau, \tag{4.16}$$

where $p.v.$ represents the Cauchy principal value. Then, the analytic representation for a real-valued signal $s(t)$ is given by

$$z(t) = s(t) + j\widetilde{s}(t) \tag{4.17}$$

The second step in the ACSP method involves the computation of the normalized complex-valued covariance matrix [92]

$$\Sigma_k = \frac{z_k z_k^*}{\text{trace}(z_k z_k^*)} \tag{4.18}$$

where $z_k \in \mathbb{C}^{n \times T}$, $k \in \{1,2\}$ denotes the analytic representation of a single EEG trial belonging to class k over N number of channels and T number of samples, and with * representing the hermitian transposition. Then, the means of the covariance matrices $\overline{\Sigma}_1$ and $\overline{\Sigma}_2$ are computed by averaging for each class the covariance matrices of the successive training trials. So that, similarly to the standard CSP approach, the complex-valued spatial filters $\mathbf{W}$ can be computed by solving the following generalized eigenvalue decomposition problem [20]

$$\overline{\Sigma}_1 \mathbf{W} = \lambda \overline{\Sigma}_2 \mathbf{W} \tag{4.19}$$

with λ denoting the diagonal matrix containing the eigenvalues of Σ_1. The spatial filtered signal $\mathbf{Y_k}$ can be obtained from the analytic representation of the EEG trials $\mathbf{Z_k}$ as

$$\mathbf{Y_k} = \mathbf{W}\mathbf{Z_k}. \tag{4.20}$$

Finally, in order to generate the ACSP features equation 4.15 can be applied as discussed in the previous section. Figure 4.7 shows a toy example where the performance of the ACSP method can be appreciated in comparison to the standard version.
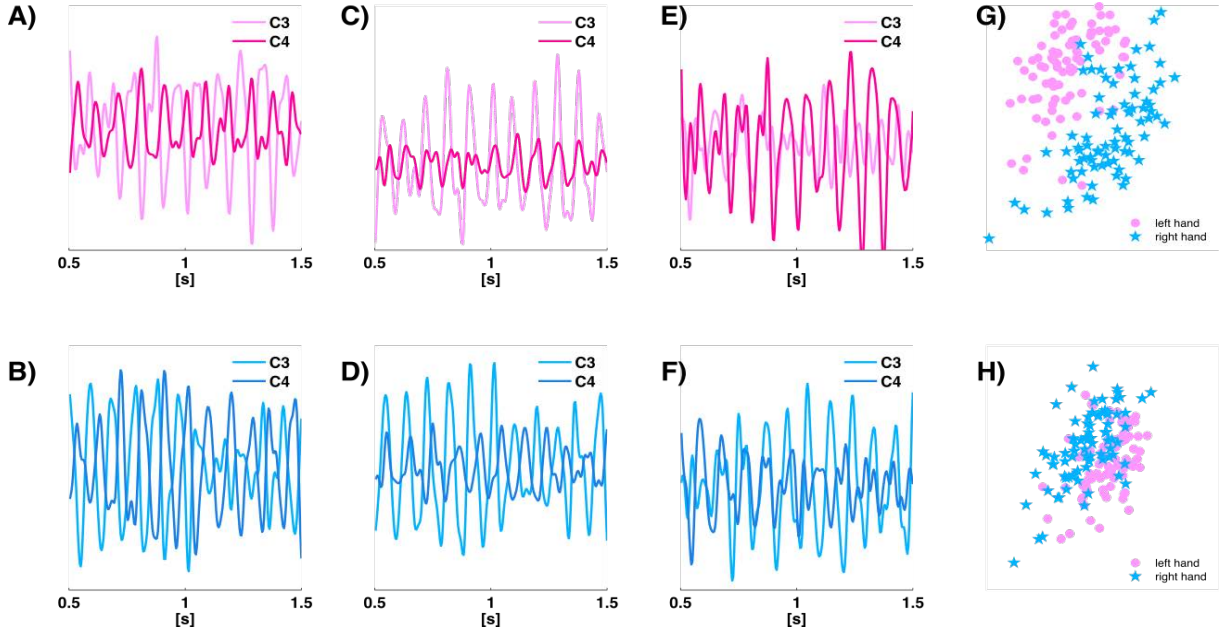
Fig. 4.7: Analytic Common Spatial Pattern algorithm. The first column of images shows the band-pass filtered EEG signals over electrodes C3 and C4 during a one-second segment of left hand motor imagery (A) and right hand motor imagery (B). In the second column the corresponding projections after applying the standard CSP algorithm over the left hand (C) and right hand (D) segments are presented. The third column shows the projections by using the analytical version of the method, namely, the ACSP algorithm over the left hand (E) and right hand (F) segments. Finally, in the last column, the obtained features after applying equation 4.15 over the projected trials by using the ACSP algorithm (G) and by using the CSP method (H) are presented. Note that separation in the case of the ACSP approach is enhanced in comparison to the standard algorithm.

4.3 Feature selection

During feature extraction many elements may be generated from several EEG channels, different time segments, multiple frequency bands, and even from many classes (e.g., under a one-versus-one approach or under a one-versus-all approach, see sections 5.1.1 and 5.1.2), before being concatenated into a single feature vector. From this situation arises one of the major difficulties in BCI design, which consists in choosing relevant elements from the broad number of possible features [2]. High dimensional feature spaces are not desirable due to the *curse of dimensionality*, which implies that the required amount of training data to obtain reliable results increases exponentially with the dimensionality of the feature space [93]. Considering the long periods of time that the training process demands, in combination with the fatigue that it causes to the users, the availability of training data is usually small in BCI research, therefore, such requirements are difficult to meet and it becomes crucial to discard part of the data, specially because in many cases, some may not be even discriminative.

The process of choosing only the most discriminative elements is known as feature selection,

which can be accomplished by examining all the possible subsets of features. However, the number of possibilities grows exponentially and, even for a moderate number of features, an exhaustive search may result impractical. Some more efficient algorithms can be applied with the aim of minimizing the number of features while maximizing the classification performance. A widely used approach consists in estimating the *mutual information* between features and the existing classes which, as explained in the next section, provides a measure of how descriptive are the involved terms to facilitate the classification decision.

4.3.1 Mutual information

The mutual information of two random variables is a measure of the mutual dependence between them. More precisely, it estimates the *amount of information* that can be obtained from one random variable through the other one. In this sense, mutual information is closely related to the concept of *entropy*, which defines the *amount of information* contained in a discrete random variable $\mathbf{x} \in \mathbb{R}^n$ in the following form

$$H(\mathbf{x}) = -\sum_{i=1}^{n} p(x_i) \log_b P(x_i), \tag{4.21}$$

where $p(x_i)$ represents the probability density function of $\mathbf{x}$ and b represents the base of the logarithm. A common choice is $b = 2$, for which equation 4.21 is referred as Shannon entropy, and whose units are expressed in bits. Considering a joint distribution $p(\mathbf{x}, \mathbf{y})$, if a value of $\mathbf{x}$ is already known, then the additional information needed to specify the corresponding value of $\mathbf{y} \in \mathbb{R}^m$ is given by

$$H(\mathbf{y}|\mathbf{x}) = -\sum_{i=1}^{n} p(x_i) \sum_{j=1}^{m} p(y_j|x_i) \log_b p(y_j|x_i), \tag{4.22}$$

which is called the *conditional entropy* of $\mathbf{y}$ given $\mathbf{x}$. Now, mutual information is a measure of the inherent dependence expressed in the joint distribution of $\mathbf{x}$ and $\mathbf{y}$ under the assumption of independence, and it is defined as

$$I(\mathbf{x}, \mathbf{y}) = H(\mathbf{x}) - H(\mathbf{y}|\mathbf{x}). \tag{4.23}$$

Thus mutual information can be understood as the reduction in the uncertainty about $\mathbf{x}$ after being told the value of $\mathbf{y}$ (or vice versa). From a Bayesian point of view, we can consider $p(\mathbf{x})$ as the prior distribution for $\mathbf{x}$ and $p(\mathbf{x}|\mathbf{y})$ as the posterior distribution after we have observed new data $\mathbf{y}$. The mutual information therefore represents the reduction in uncertainty about $\mathbf{x}$ as a consequence of the new observation $\mathbf{y}$ [87].

In the context of feature selection, a measure of mutual information reveals how much information the presence/absence of a term contributes to making the correct classification decision.

4.4 Filter Bank Common Spatial Pattern (FBCSP)

The effective use of the CSP algorithm discussed in section 4.1.3 requires the assignment of certain parameters, namely, the frequency range for the band-pass filtering of the EEG signals, the time interval relative to the task cue for extracting the trials, and the number of CSP filters pairs to be used [20]. These parameters are generally fixed by a standard systematization of the method, so that the frequency range involves the alpha and beta bands (e.g., within the range from 8 to 30 Hz), the time segment is taken after 0.5 or 1 seconds after the stimuli, and 2 or 3 pairs of filters are considered. However, the performance of the CSP algorithm can be potentially enhanced if subject-specific parameters are taken into account [94]. In this sense, the Filter Bank Common Spatial Pattern (FBCSP) algorithm enhances the performance of the original CSP method by automatically selecting the most discriminative subject-specific frequency range for band-pass filtering of EEG measurements [95].

The FBCSP includes four progressive stages of signal processing and machine learning that are applied on the EEG data:

1. A filter bank comprising multiple band-pass filters, which decomposes the EEG signals into multiple frequency pass bands. Various configurations of the filter bank are effective, however, a good choice consists in covering the frequency range from 4 to 40 Hz by using 9 filters of 4 Hz width, namely, 4-8, 8-12,..., 36-40 Hz. This selection allows producing stable filters' responses while covering the physiological range of frequencies that is of interest for EEG analysis.

2. Spatial filtering in terms of equation 4.14 over the 9 frequency ranges of the EEG trials generated in the previous step.

$$\mathbf{Y}_{b,i} = \mathbf{W}_b^\top \mathbf{E}_{b,i},$$

where $\mathbf{E}_{b,i} \in \mathbb{R}^{c \times t}$ represents the single-trial EEG measurement from the bth band-pass filter of the ith trial; $\mathbf{Y}_{b,i} \in \mathbb{R}^{c \times t}$ denotes $\mathbf{E}_{b,i}$ after spatial filtering, $W_b \in \mathbb{R}^{c \times c}$ corresponds to the CSP filters coefficient matrix; c is the number of channels; t is the number of EEG samples per channel; and $^\top$ denotes the transpose operator. The m pairs of CSP features of the ith trial for the bth band-pass filtered EEG measurements are computed in terms of equation 4.15 in the following terms

$$\mathbf{v}_{b,i} = \log \frac{\text{diag}(\widetilde{\mathbf{W}}_b^\top \mathbf{E}_{b,i} \mathbf{E}_{b,i}^\top \widetilde{\mathbf{W}}_b)}{\text{trace}[\widetilde{\mathbf{W}}_b^\top \mathbf{E}_{b,i} \mathbf{E}_{b,i}^\top \widetilde{\mathbf{W}}_b]}$$

where $\mathbf{v}_{b,i} \in \mathbb{R}^{2m}$; $\widetilde{\mathbf{W}}_b$ represents the first m and the last m columns of W_b; diag($\cdot$) returns the diagonal elements of the square matrix; and trace[$\cdot$] returns the sum of the diagonal

elements in the square matrix. (Note that $m = 2$). The resulting feature vector for the ith trial is then arranged as follows

$$\mathbf{v}_i = [\mathbf{v}_{1,i}, \mathbf{v}_{2,i}, ..., \mathbf{v}_{9,i},],$$

where $\mathbf{v}_i \in \mathbb{R}^{1 \times (9*2)}, i = 1, 2, ..., n$; with n denoting the total number of EEG trials.

3. Feature selection to extract the most discriminative CSP features from $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n]$. To this end several algorithms can be used, however, in the present work we have decided to apply the Mutual Information-based Best Individual Feature (MIBIF) [90], [95], which is described in the following section.

4. Classification to translate the generated and selected features. For this task there are also many possible choices, in the present work we will explore different solutions by using linear discriminant analysis and support vectors machines. Both techniques are discussed later along this chapter.

4.4.1 Mutual Information-based Best Individual Feature (MIBIF) algorithm

In brief, this algorithm consists in computing the mutual information of the considered features and sorting the results in a descending order, so that only k features are selected. For a 2-class problem, the MIBIF starts by computing the mutual information in the following way

$$H(f_j; w) = H(w) - H(w|f_j), \tag{4.24}$$

with $w = 1, 2 \in C$ representing the class labels, and the terms $H(w)$ and $H(w|f_j)$ consisting respectively of the entropy and conditional entropy, both defined as

$$H(w) = -\sum_{w=1}^{2} p(w) \log_2 p(w), \tag{4.25}$$

$$H(w|f_j) = -\sum_{w=1}^{2} P(w|f_j) \log_2 P(w|f_j) \tag{4.26}$$

$$= -\sum_{w=1}^{2} \sum_{i=1}^{n} p(w|f_{j,i}) \log_2 p(w|f_{j,i}), \tag{4.27}$$

where the feature value of the ith trial from f_j is represented by $f_{j,i}$. Now, the probability $p(w|f_{j,i})$ can be computed by using the Bayes theorem as

$$p(w|f_{j,i}) = \frac{p(f_{j,i}|w)P(w)}{p(f_{j,i})}, \tag{4.28}$$

where $p(w|f_{j,i})$ is the conditional probability of class w given $f_{j,i}$; $p(f_{j,i}|w)$ is the conditional probability of $f_{j,i}$ given class w; $P(w)$ is the prior probability of class w; and $p(f_{j,i})$ is given by

$$p(f_{j,i}) = \sum_{w=1}^{2} P(f_{j,i}|w)P(w). \tag{4.29}$$

The conditional probability $p(f_{j,i}|w)$ can be estimated by using a Parzen window [96] in the following terms

$$\hat{p}(f_{j,i}|w) = \frac{1}{n_w} \sum_{k \in I_w} \phi(f_{j,i} - f_{j,k}, h), \tag{4.30}$$

where n_w is the number of data samples belonging to class w; I_w is the set of indexes of the training data trials belonging to class w; $f_{j,k}$ is the feature value of the kth trial from f_j and ϕ is a smoothing kernel function with a smoothing parameter h given by

$$\phi(y, h) = \frac{1}{\sqrt{2\pi}} exp - \left(\frac{y^2}{2h^2} \right), \tag{4.31}$$

where

$$h^{opt} = \left(\frac{4}{3n} \right)^{\frac{1}{5}} \sigma. \tag{4.32}$$

By substituting equations 4.30 and 4.31 into 4.28 we can proceed to compute the mutual information associated to each feature, from which the k elements presenting the highest values will be selected. For the implementation of the FBCSP algorithm we have decided to use $k = 4$. Note that, since the CSP features are paired, this number fluctuates between 4 components (if the selected elements come from the same 2 pairs) to 8 (if each one of the elements belongs to a different pair).

4.5 Common Spatial Pattern by Joint Approximate Diagonalization (CSP by JAD)

Classical CSP extensions for solving multiclass problems are carried out by performing a series of binary CSP (e.g., by computing CSP under a one-vs-one approach or under a one-vs all approach). Another possible technique consists in computing the joint approximate diagonalization (JAD) over the mean covariance matrices of the involved classes, which is the same principle of the eigenvalue decomposition performed by the standard CSP method for two classes [15].

Given EEG data recorded over N number of channels and T number of time samples from K different classes, the strategy of the CSP by JAD method consists in finding a transformation

$\boldsymbol{W} \in \mathbb{R}^{N \times N}$ that diagonalizes the covariance matrices $\boldsymbol{R}_x|c_i$, so that

$$\boldsymbol{W} \boldsymbol{R}_{x|c_i} \boldsymbol{R}_{x|c_i} \boldsymbol{W}^\top = \boldsymbol{D}_{c_i}, i = 1, ..., K, \tag{4.33}$$

with $\boldsymbol{D}_{c_i} \in \mathbb{R}^{N \times N}$ diagonal matrices, and with $\boldsymbol{W}$ representing the spatial filters coefficient matrix. An effective approach to generate the transformation matrix $\boldsymbol{W}$ consists of applying the Fast Frobenius Diagonalization (FFDIAG) [97].

Now, each column vector of $\boldsymbol{W}$ represents a spatial filter, however, not all of them are relevant for the subsequent classification, so that only the most discriminative columns should be selected. To this end, a preprocessing scheme based on information theoretic feature extraction (ITFE) can be implemented, from which only the L elements generating the highest mutual information values are considered as the optimal filters [98]. The general framework of ITFE and its application to solve the aforementioned problem is presented in the following section.

4.5.1 Information theoretic feature extraction (ITFE)

In the context of non-invasive BCIs, the goal is to find a dimension-reduced representation of EEG data that maximizes the mutual information of class labels and extracted EEG components. We thus have $\mathcal{X} = \mathbb{R}^{T \times N}$, and choose $\hat{\mathcal{X}} = \mathbb{R}^T$ for the dimension-reduced feature space. In this sense, EEG components are sequentially extracted. Now, let $\mathbf{x} \in \mathcal{X}$ be a random variable (e.g., the recorded EEG data), from which $c \in \mathcal{C}$ (e.g., the class label to be predicted). Then, the general idea of ITFE, is to find a transformation $f^*: \mathbf{x} \to \hat{\mathcal{X}}$ that maps the original feature space $\mathcal{X}$ into a discrete set $\hat{\mathcal{X}}$ while preserving information on the class labels c in $\mathbf{x}$

$$f* = \underset{f \in F}{\operatorname{argmax}}\{I(c, f(\mathbf{x}))\} \tag{4.34}$$

with $I(c, f(\mathbf{x}))$ the mutual information of c and $f(\mathbf{x})$, and $\mathcal{F}$ a function space. If we consider only linear transformations, equation 4.34 thus simplifies to

$$w* = \underset{\mathbf{w} \in R^N}{\operatorname{argmax}}\{I(c, \mathbf{w}^\top \mathbf{x}\} \tag{4.35}$$

Now, note that the mutual information of c and $\hat{x} = \mathbf{w}^\top \mathbf{x}$ can be written as (see section 4.3.1)

$$I(c, \mathbf{w}^\top \mathbf{x}) = H(\mathbf{w}^\top \mathbf{x}) - H((\mathbf{w}^\top \mathbf{x})|c) = H(\hat{x}) - H(\hat{x}|c)$$
$$= H(\hat{x}) - \sum_{i=1}^{K} P(c_i)H(\hat{x}|c). \tag{4.36}$$

If we consider $p(\mathbf{x}|c) = \mathcal{N}(0, R(\mathbf{x}|c))$, from where it follows that $p(\hat{x}|c) = \mathcal{N}(0, R(\hat{x}|c))$, and we assume $\sigma_{\hat{x}}^2 = 1$, which does not imply loss of generality given that $\mathbf{w}$ can always be scaled to meet this assumption, we obtain an estimate of the mutual information of c and $\hat{x}$ in the following

terms [98]

$$I(c, \mathbf{w}_j^\top \mathbf{x}) \approx - \sum_{i=1}^{K} P(c_i) \log \sqrt{\mathbf{w}_j^\top \boldsymbol{R}_{\mathbf{x}|c_i} \mathbf{w}_j} - \frac{3}{16} \Big(\sum_{i=1}^{K} P(c_i)((\mathbf{w}_j^\top \boldsymbol{R}_{\mathbf{x}|c_i} \mathbf{w}_j)^2 - 1) \Big)^2 \quad (4.37)$$

The complete procedure of the multiclass ITFE is summarized in algorithm 1. Note that the output of the method corresponds with the L columns of $\boldsymbol{w}$ generating the highest mutual information. In [98] this number was arbitrarily set to $L = 8$, from which reliable results where obtained. This parameter can also be chosen by using automatic methods [99]. However, this aspect is excluded from the scope of the present work. Thus, in order to evaluate the performance of the CSP by JAD algorithm in comparison to the efficiency attained by the other considered methods, we will use the recommended value of $L = 8$. Once the L filters have been selected, feature extraction can be accomplished in terms of equation 4.15, so that discriminative elements reflecting the contrast between the variance of the projected EEG trials are generated.

Figure 4.8 shows the classification space generated by the CSP by JAD algorithm for a 3-class problem involving EEG signals during left hand, right hand, and both hands motor imageries of a right-handed male subject that has participated in the development of the 4-*class* database (see section 2.6.1). EEG signals from the entire set of channels were band-pass filtered within the frequency range from 8 to 30 Hz, and features were extracted by using equation 4.15 over the selected $\mathbf{w}_j$ filters. For visualization purposes, we show only the EEG features calculated from the 2 filters generating the highest mutual information, so that the resulting vectors are defined within a 2-dimensional space. Note that elements belonging to different conditions are well separated, however, by considering a higher number of filters their separation can be improved.

Algorithm 1 Multiclass Information Theoretic Feature Extraction (ITFE)

Input: Covariance matrices $\mathbf{R}_x|c_i$, $i = 1, ..., K$.
   1) Compute joint approximate diagonalization, so that $\mathbf{wR}_{\mathbf{x}|c_i}\mathbf{R}_{\mathbf{x}|c_i}\mathbf{w}^\top = \mathbf{D}_{c_i}, i = 1, ..., K$
     (e.g., with the FFDiag-algorithm [97]).
   2) For each column $\mathbf{w}_j$, $j = 1, ..., N$ of $\mathbf{w}$ scale $\mathbf{w}_j$, so that $\mathbf{w}_j^\top \mathbf{R}_{\mathbf{x}} \mathbf{w}_j = 1$, and estimate mutual information according to

$$I(c, \mathbf{w}_j^\top \mathbf{x}) \approx - \sum_{i=1}^{K} P(c_i) \log \sqrt{\mathbf{w}_j^\top \mathbf{R}_{\mathbf{x}|c_i} \mathbf{w}_j} - \frac{3}{16} \Big(\sum_{i=1}^{K} P(c_i)((\mathbf{w}_j^\top \mathbf{R}_{\mathbf{x}|c_i} \mathbf{w}_j)^2 - 1) \Big)^2$$

   3) Select the L columns of $\mathbf{w}$ with highest mutual information.
Output: Spatial filter coefficient matrix $\mathbf{w} \in \mathbb{R}^{N \times L}$.
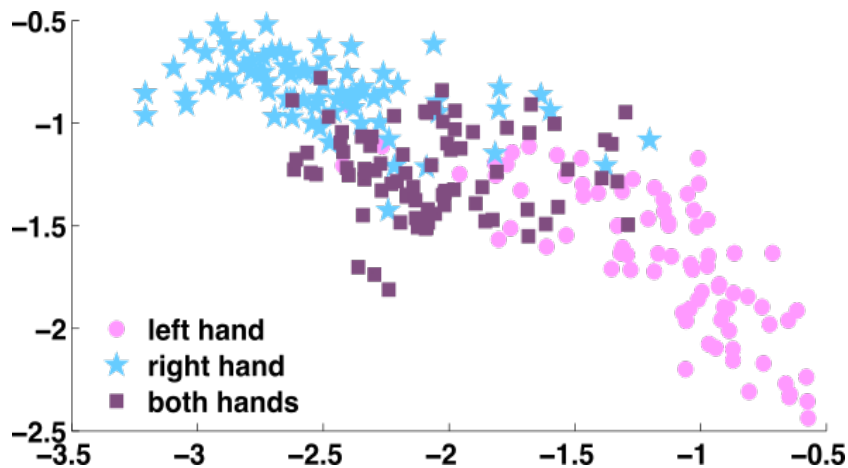
Fig. 4.8: Classification space generated by the CSP by JAD algorithm for a 3-class problem including left hand, right hand, and both hands motor imageries. EEG signals from 26 channels were band-pass filtered within the frequency range from 8 to 30 Hz, and a 4-second window starting 0.5 s after the task cue initiation was considered for the feature extraction procedure. Note that only 2 filters were selected in order to generate a 2-dimensional representation and, even though classes are already well-separated, performance of the CSP by JAD algorithm is expected to improve with a higher number of selected filters.

4.6 Classification

The aim of classification is to assign a given input vector $\mathbf{x}$ to one of K discrete classes C_k, where $k = 1,..., K$. In an ideal case, classes are considered to be disjoint, so that each input is assigned to one and only one class. The input space, or classification space, is then divided into different regions (e.g, each region associated to each one of the K classes), where boundaries are called *decision boundaries* or *decision surfaces* [87].

4.6.1 Discriminant functions

In the context of classification, a discriminant function defines a decision boundary that separates the input space into K different regions associated to K classes denoted C_k. The evaluation of an input vector $\mathbf{x}$ reflects its position with respect to the decision boundary, so that it can be associated to a specific region, and hence to the corresponding class. In this section we address attention to linear discriminants, namely those for which such a function defines an hyperplane in the following terms

$$y(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + w_0. \tag{4.38}$$

The evaluation of $y(\mathbf{x})$ generates a real number, however, the goal is to predict discrete class labels. To this end, we can consider the sign generated after its evaluation, which determines the position of the input vector with respect to the decision boundary, given that any point generating a negative value will be found below the hyperplane, whereas a positive one will be

situated above of it. With this in mind, we can consider a generalization of this model in which the linear function is transformed by using $f(\cdot)$ so that

$$y(\mathbf{x}) = f(\mathbf{w}^\top \mathbf{x} + w_0). \qquad (4.39)$$

In this context $f(\cdot)$ is known as the *activation function*, and can be defined as

$$f(y(\mathbf{x})) = \begin{cases} C_1, & \text{if } y(\mathbf{x}) < 0 \\ C_2, & \text{if } y(\mathbf{x}) > 0 \end{cases}, \qquad (4.40)$$

where the two regions represented by the decision surface have been associated with classes C_1 and C_2, so that any point located below the decision boundary belongs to C_1 and, any point located above it, belongs to C_2. Note that the corresponding decision boundary is defined by the relation $y(\mathbf{x}) = 0$, the assignment of any vector situated on the decision surface to a given class depends on the context of the classification problem.

Even if the function $f(\cdot)$ is nonlinear, the decision surfaces are linear functions of $\mathbf{x}$, therefore, the class of models described by equation 4.40 are called *generalized linear models* [100]. When classes can be completely separated by linear decision boundaries, the corresponding data set is said to be *linearly separable*.

To facilitate the discussion, we start with the case of two classes and then we consider the extension to $K > 2$.

4.6.1.1   Two classes

As mentioned, the simplest model of a linear discriminant function is given by a linear function of the input vector so that

$$y(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + w_0,$$

where the parameters $\mathbf{w}$ and w_0 represent, respectively, the *weight vector* and *bias* of the resulting hyperplane. If the evaluation of $y(\mathbf{x}) < 0$, the input vector $\mathbf{x}$ is assigned to class C_1, otherwise, to class C_2. The corresponding decision boundary is therefore defined by the relation $y(\mathbf{x}) = 0$, which defines a $(D-1)-$dimensional hyperplane within the $D-$dimensional input space. Let us consider two points, $\mathbf{x}_1$ and $\mathbf{x}_2$, both of them lying on the decision surface. Given that $y(\mathbf{x}_1) = y(\mathbf{x}_2) = 0$, it follows that $\mathbf{w}^\top(\mathbf{x}_1 - \mathbf{x}_2) = 0$ and hence the vector $\mathbf{w}$ is orthogonal to every vector located on the decision surface, therefore, $\mathbf{w}$ determines the orientation of the hyperplane. In the same way, if $\mathbf{x}$ is a point lying within the decision surface, then $y(\mathbf{x}) = 0$, so that the normal distance from the origin to the decision surface corresponds to

$$\frac{\mathbf{w}^\top \mathbf{x}}{\|\mathbf{w}\|} = -\frac{w_0}{\|\mathbf{w}\|}. \qquad (4.41)$$

Based on this, we can see that w_0 defines the location of the decision boundary. These

properties are represented in figure 4.9 for a 2−dimensional case.

Moreover, the evaluation of $y(\mathbf{x})$ generates a signed measure of the perpendicular distance r of the point $\mathbf{x}$ from the decision boundary. To observe this, consider an arbitrary point $\mathbf{x}$ and let $\mathbf{x}_\perp$ be its orthogonal projection onto the decision surface, so that

$$\mathbf{x} = \mathbf{x}_\perp + r\frac{\mathbf{w}}{\|\mathbf{w}\|}. \tag{4.42}$$

By multiplying both sides of this expression by $\mathbf{w}^\top$ and adding w_0, we can make use of $y(\mathbf{x}) = \mathbf{w}^\top\mathbf{x} + w_0$ and $y(\mathbf{x}_\perp) = \mathbf{w}^\top\mathbf{x}_\perp + w_0 = 0$ to observe that (see fig. 4.9)

$$r = \frac{y(\mathbf{x})}{\|\mathbf{w}\|}. \tag{4.43}$$
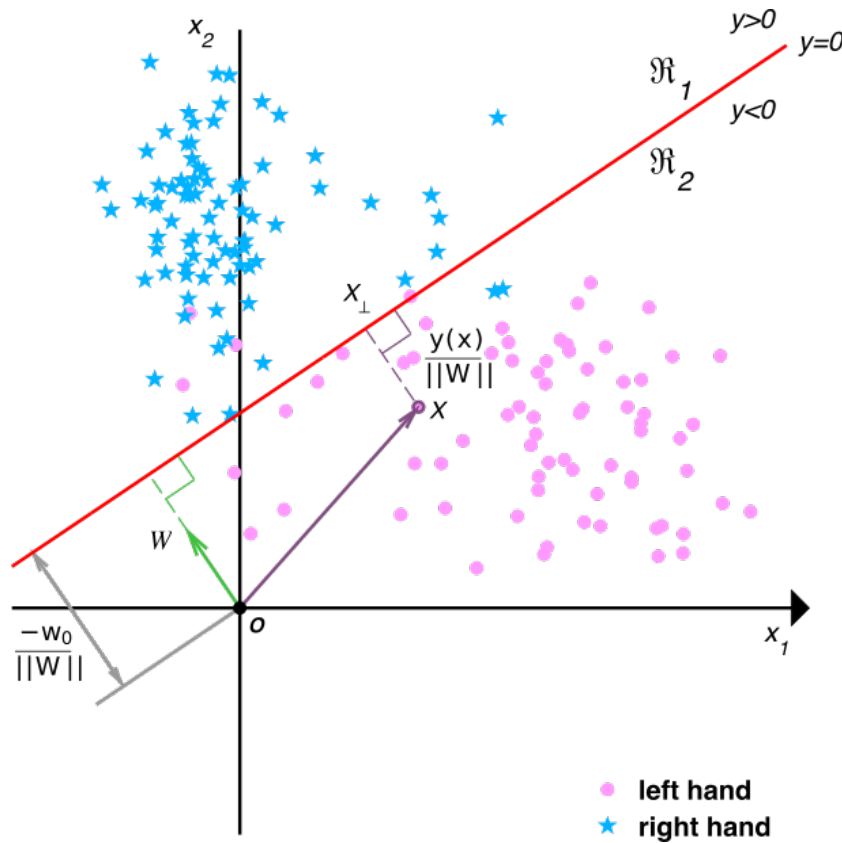
Fig. 4.9: Geometry of a linear discriminant function in two dimensions generated by LDA. The decision boundary is shown in red, which is perpendicular to vector $\mathbf{w}$, and presents a separation from the origin that is controlled by the bias parameter w_0. The signed orthogonal distance of a given point $\mathbf{x}$ from the decision boundary is obtained as $y(\mathbf{x})/\|\mathbf{w}\|$. Data points correspond with the features extracted by the CSP algorithm presented in section 4.1.3, which were obtained in order to discriminate between EEG signals during left hand and right hand motor imageries
.

4.6.1.2 Multiple classes

Let us consider the extension of linear discriminants for cases involving more than two classes ($K > 2$). Intuitively, we might think about building a K-class discriminant by combining a number of two-class discriminant functions. For instance, consider the use of $K(K-1)/2$ binary discriminant functions, each one trained in order to separate every pair of classes. This approach is known as a one-versus-one classifier (see section 5.1.1), and consists in assigning the class label C_k to each input vector according to a majority vote amongst the discriminant functions. However, this approach leads to ambiguous regions, as exhibited in figure 4.10-A, where the class assignment is compromised.

We can overcome these difficulties by considering a single K-class discriminant including K linear functions of the form

$$y_k(\mathbf{x}) = \mathbf{w}_k^\top \mathbf{x} + w_{k0}, \tag{4.44}$$

so that a point $\mathbf{x}$ is assigned to class C_k if $y_k(\mathbf{x}) > y_j(\mathbf{x}), \forall j \neq k$, or to class C_j otherwise. The decision boundary between class C_k and class C_j is therefore given by $y_k(\mathbf{x}) = y_j(\mathbf{x})$, which can be defined as

$$(\mathbf{w}_k - \mathbf{w}_j)^\top \mathbf{x} + (w_{k0} - w_{j0}) = 0. \tag{4.45}$$

This model corresponds to a $(D-1)$-dimensional hyperplane whose decision regions are always singly connected and convex (see fig. 4.10-B).

Equation 4.45 has the same form as the one discussed in section 4.6.1 for finding the decision boundary in a two-class case, and hence analogous geometrical properties apply. In fact, this formalism is also pertinent for solving a binary problem, from which the decision boundary can be defined based on two discriminant functions $y_1(\mathbf{x})$ and $y_2(\mathbf{x})$. However, the formulation based on a single discriminant function $y(\mathbf{x})$ suggests a simpler approach.

In the next two sections we will explore two approaches to learning the parameters of linear discriminant functions, namely, Linear Discriminant Analysis (LDA) and Support Vector Machines (SVMs).

4.6.1.3 Linear Discriminant Analysis (LDA)

The linear discriminant analysis for classification can be motivated by using a Bayesian approach, from which the probability of an input vector $\mathbf{x}$ belonging to class C_k can be estimated in the following terms

$$p(C_k|\mathbf{x}) = \frac{p(\mathbf{x}|C_k)p(C_k)}{p(\mathbf{x})}. \tag{4.46}$$

Now, if we want to determine to which class $\mathbf{x}$ belongs to, it is possible to apply the *Bayes decision rule*, which implies that
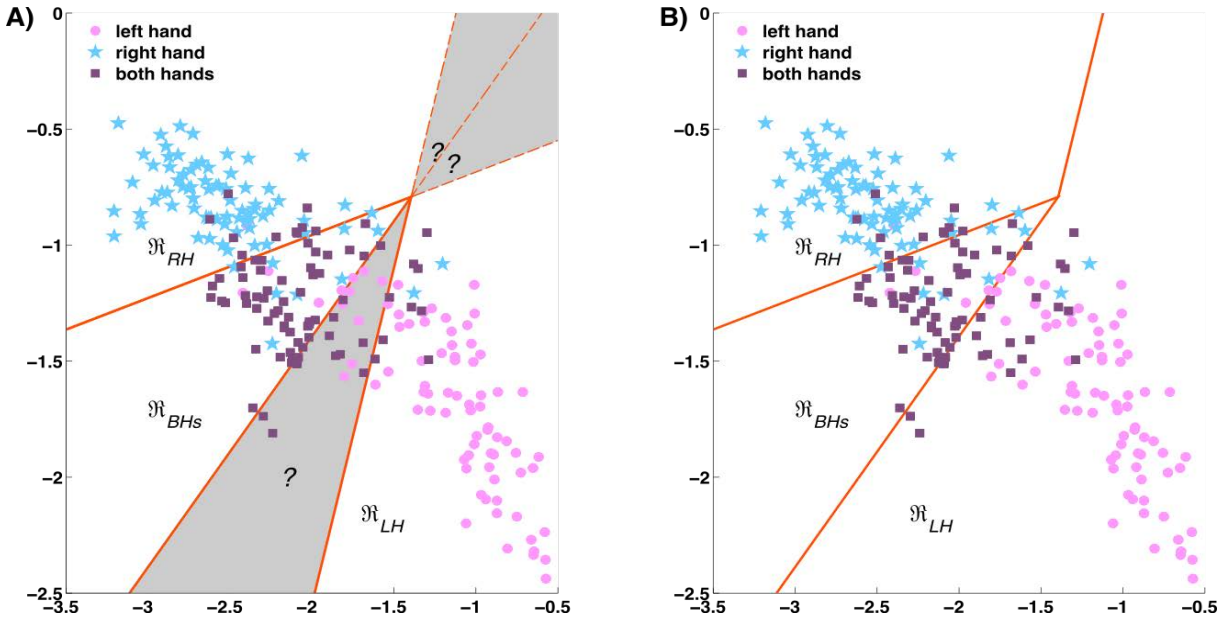
Fig. 4.10: Discriminant functions for multiple classes. Attempting to construct a K class discriminant from a set of two class discriminants leads to ambiguous regions, shown in the left side. This problem can be overcome by considering the solution expressed in equation 4.45, which leads to decision boundaries that are singly connected and convex.

$$\text{Decide } C_1 \text{ if } p(C_1|\mathbf{x}) > p(C_2|\mathbf{x}); \text{ otherwise decide } C_2.$$

The generalization of such decision rule to multiple classes is straightforward and it can be rewritten as

$$
\begin{aligned}
g(\mathbf{x}) &= \max_{k=1,...,K}(p(\mathbf{x}|C_k)p(C_k)) \\
&= \max_{k=1,...,K}(\log(p(\mathbf{x}|C_k)p(C_k))).
\end{aligned}
\tag{4.47}
$$

Let us consider the case of multivariate normal distributions, for which the term $p(\mathbf{x}|C_k)p(C_k)$ is given by equation 4.2, and let us assume that all distributions have the same covariance matrix, then, equation 4.47 derives to

$$
g(\mathbf{x}) = \max_{k=1,...,K}\left(-\frac{1}{2}(\mathbf{x} - \mu_k)^\top \Sigma^{-1}(\mathbf{x} - \mu_k) + \log(p(C_k) + \delta)\right).
\tag{4.48}
$$

The first term in equation 4.48 is the square form of the Mahalanobis distance (see section 4.1.2), which correspond with the separation from the input data to the corresponding class mean μ_k, $p(C_k)$ is the prior probability of class k, and δ is a constant involving all terms that do not

depend on k, and hence can be removed, which leads to

$$g(\mathbf{x}) = \max_{k=1,\dots,K} (\mathbf{x}^\top \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^\top \Sigma^{-1} + \log(p(C_k))). \tag{4.49}$$

Now, we can define the linear discriminant function as

$$y_k(\mathbf{x}) = \mathbf{x}^\top \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^\top \Sigma^{-1} + \log(p(C_k)), \tag{4.50}$$

then,

$$g(\mathbf{x}) = \max_{k=1,\dots,K} (y_k(\mathbf{x})). \tag{4.51}$$

For a two class problem, the decision boundary between $k = 1$ and $k = 2$ corresponds with

$$y_1(\mathbf{x}) = y_2(\mathbf{x}), \tag{4.52}$$

or equivalently the following holds

$$\mathbf{x}^\top \Sigma^{-1}(\mu_1 - \mu_2) - \frac{1}{2}(\mu_1 + \mu_2)^\top \Sigma^{-1}(\mu_1 - \mu_2) + \log \frac{p(C_{k=1})}{p(C_{k=2})} = 0. \tag{4.53}$$

Equation 4.53 can be interpreted as the linear model representing the hyperplane of the decision boundary discussed in section 4.6.1, so that

$$y(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + w_0, \quad \text{with} \quad \begin{cases} \mathbf{w}^\top = \Sigma^{-1}(\mu_1 - \mu_2) \\ w_0 = \frac{1}{2}(\mu_1 + \mu_2)^\top \Sigma^{-1}(\mu_1 - \mu_2) + \log \frac{p(C_{k=1})}{p(C_{k=2})} \end{cases} \tag{4.54}$$

The parameters Σ, μ and $p(k)$ are obtained from the training data set. If we consider a specific class covariance Σ_k for each class, then the classifier is no longer linear and is referred to as quadratic discriminant analysis (QDA).

Figure 4.9 shows the decision boundary obtained through linear discriminant analysis for the features extracted by the CSP algorithm presented in section 4.1.3, which were obtained in order to discriminate between EEG signals during left hand and right hand motor imageries.

4.6.1.4 Support Vector Machines (SVM)

A support vector machine is a supervised classification algorithm based on the statistical learning theory developed by Vladimir Vapnik back in the 1970s [101]. Their functionality is based on a binary task, for which they aim at finding the hyperplane that maximizes the separation between the closest data belonging to different conditions (i.e., support vectors). This separation is called the *margin*, and it can be enhanced by mapping the corresponding features into a new space of higher dimension. To this end, they apply a kernel function that transforms the input space into

a new one, where discrimination of classes from different conditions is reinforced.

To start with, let us consider the function defining an hyperplane given by equation 4.38

$$y(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + w_0,$$

Now, the goal of SVMs is to find the hyperplane that maximizes the margin between the selected support vectors. Such a margin is set so that the hyperplane lies equidistant from the support vectors belonging to different classes with a geometric distance of 1, as it is shown in figure 4.11. Then, the geometric distance for any input vector will satisfy

$$d(\mathbf{x}, \mathbf{w}, b) = \frac{|\mathbf{w}^\top \mathbf{x} + b|}{\|\mathbf{w}\|} \geq \frac{1}{\|\mathbf{w}\|}, \tag{4.55}$$

so that $|\mathbf{w}^\top \mathbf{x} + b| = 1$ for the support vectors and $|\mathbf{w}^\top \mathbf{x} + b| > 1$ for the other training points (non-support vectors). This implies that the margin of the hyperplane ρ can be defined as twice the distance separating the support vectors from the hyperplane, so that

$$\rho = \frac{2}{\|\mathbf{w}\|}. \tag{4.56}$$

In this way, the classification algorithm is able to find the optimal solution. Although, there is an infinite number of hyperplanes that solve a linearly separable problem, there is only one that maximizes the margin ρ in such a way that the parallel hyperplanes to the decision boundary that are drawn on the support vectors satisfy $\mathbf{w}^\top \mathbf{x} + b = 1$ and $\mathbf{w}^\top \mathbf{x} + b = -1$ (see fig. 4.11). The parameters $\mathbf{w}$ and b that hold this assumption are found during the training stage. The goal is to maximize the margin distance ρ, which can be achieved by minimizing $\|\mathbf{w}\|$, or in order to facilitate the task without loss of generality, by minimizing $\frac{1}{2}\|\mathbf{w}\|^2$ in the following terms

$$J(\mathbf{w}) = \min\left(\frac{1}{2}\|\mathbf{w}\|^2\right), \quad \text{s.t.,} \ |\mathbf{w}^\top \mathbf{x} + b| \geq 1, \quad \forall i \in \mathcal{X}, \tag{4.57}$$

with $\mathcal{X}$ representing the training data set indexes. In most of the cases, problems are not completely linearly separable, so that it is not possible to meet this criterion. In order to accept a more flexible solution, the constraint in equation 4.57 is replaced by

$$\mathbf{w}^\top \mathbf{x} + b \geq 1 - \xi_i, \quad \forall i \in \mathcal{X}, \tag{4.58}$$

where the term ξ_i represents the so-called *slack* variables, which allow support vectors to be in the margin if the evaluation of equation 4.58 complies with $0 \leq \xi_i \leq 1$. To avoid having a large number of misclassified data, the penalty term P is added to $J(\mathbf{w})$, so that

$$J(\mathbf{w}) = \min\left(\frac{1}{2}\|\mathbf{w}\|^2\right) + P\sum_{i \in \mathcal{X}} \xi_i, \quad \text{s.t.,} \ y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i \in \mathcal{X} \tag{4.59}$$

The compromise between maximizing the margin and minimizing the amount of slack is given by the constant P, also referred to as *box constraint*. There are two different approaches for solving this problem, on the one hand, $J(\mathbf{w})$ can be directly addressed upon the primal optimization problem [102]. And, on the other hand, it is possible to use *Lagrange multipliers* for solving the dual optimization problem, which actually is the most common technique [103]. By using the second alternative, the optimization problem is expressed in terms of α_i in the following terms

$$D(\alpha) = \max \Big(\sum_{i \in \mathcal{X}} \alpha_i - \frac{1}{2} \sum_{i \in \mathcal{X}} \sum_{j \in \mathcal{X}} y_i y_j \alpha_i \alpha_j \mathbf{x}_i \mathbf{x}_j \Big), \quad \text{s.t.,} \quad \sum_{i \in \mathcal{X}} \alpha_i y_i = 0 \quad 0 \geq \alpha_i \leq P, \quad \forall i \in \mathcal{X}.$$

(4.60)

Once α has been found, the optimal parameters are given by

$$y(\mathbf{x}) = \sum_{i \in \mathcal{X}} \alpha_i y_i \mathbf{x}_i + b.$$

(4.61)

Only the support vectors $\mathbf{x}_i$, which define the limits of the margin and the separating hyperplane, meet the condition $\alpha_i > 0$.

As mentioned, when the elements from different classes are not linearly separable, it is possible to extend the SVM algorithm to a non-linear approach by implicitly mapping data into a higher-dimensional feature space, where classes are linearly separable. This solution is achieved by applying the so-called *kernel trick* [104], which uses a kernel function $\mathcal{K} = \varphi(\mathbf{x}_i)\varphi(\mathbf{x}_j)$ to define the new feature space, with $\varphi(\cdot)$ representing the transformation over the input data. When such a kernel function is applied, equation 4.60 takes the following form

$$D(\alpha) = \max \Big(\sum_{i \in \mathcal{X}} \alpha_i - \frac{1}{2} \sum_{i \in \mathcal{X}} \sum_{j \in \mathcal{X}} y_i y_j \alpha_i \alpha_j \mathcal{K}(\mathbf{x}_i \mathbf{x}_j) \Big), \quad \text{s.t.,} \quad \sum_{i \in \mathcal{X}} \alpha_i y_i = 0 \quad 0 \leq \alpha_i \leq P, \quad \forall i \in \mathcal{X}.$$

(4.62)

The most common kernel functions are linear, polynomial and radial basis, which are defined as

— Linear: $\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \mathbf{x}_j$

— Polynomial: $\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i^\top \mathbf{x}_j + r)^d, \quad \gamma > 0$

— Radial basis function: $\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) = e^{(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)}, \quad \gamma > 0$

The original SVM model is not directly applicable to classification problems involving more than two classes. However, there are several approaches for its multiclass extension, such as the

one-versus-one and one-versus-all techniques which will be both explained in chapter 5.
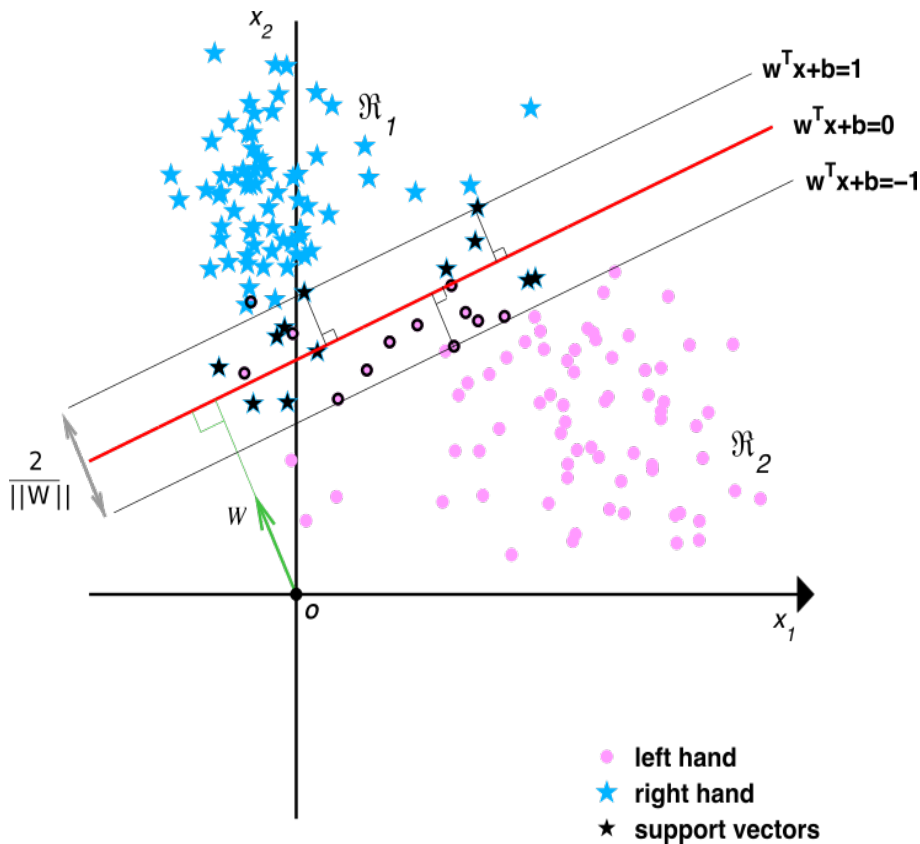
Fig. 4.11: Support Vectors Machines classification by illustrating the margin maximization.
.

4.6.2 Distance-based classification

In order to explore different solutions to the CSP algorithm, we will present a small outlook to an alternative approach that is known as distance-based classification. The idea of these approaches is to assign the class label of an unknown element according to the shortest distance that separates it from the class measures defined during the training phase. This approach has drawn special attention due to the possibility of applying it in the context of Riemannian geometry, which will be elucidated in the following section.

4.6.2.1 Riemannian geometry

Square real symmetric positive-definite (SPD) matrices have the property of being always diagonalizable with strictly real positive eigenvalues, whose space represent a differentiable Riemannian manifold $\mathcal{M}$. Using this manifold offers several advantages that can be easily applied in the context of signal processing [13], [105]. Note that the covariance matrices are SPD, so that they belong to a manifold representing a Riemannian space from where their classification can be

directly achieved.

Let denote by $S(n) = \{\mathbf{S} \in M(n), S^\top = \mathbf{S}\}$ the space of all $n \times n$ symmetric matrices in the space of square real matrices $M(n)$; and $P(n) = \{\mathbf{P} \in S(n), \mathbf{u}^\top \mathbf{P}\mathbf{u} > 0, \forall\ \mathbf{u} \in \mathbb{R}^n\}$ the set of all $n \times n$ SPD matrices. The derivatives at matrix $\mathbf{P}$ on the manifold lie in a vector space T_p, which is the tangent space at that point (Figure 4.12). Now, let $\mathbf{\Gamma}(t): \to P(N)$ be any differentiable path from $\Gamma(0) = \mathbf{P}_1$ to $\Gamma(0) = \mathbf{P}_2$. The length of $\mathbf{\Gamma}(t)$ is given by

$$L(\mathbf{\Gamma}(t)) = \int_0^1 \|\dot{\mathbf{\Gamma}}\|_{\mathbf{\Gamma}(t)}, dt. \tag{4.63}$$

The minimum length curve connecting two points in the manifold is called the geodesic, and it gives the Riemannian distance between the two points

$$\delta_R(\mathbf{P}_1, \mathbf{P}_2) = \| \log(\mathbf{P}_1^{-1}\mathbf{P}_2) \|\ _F = \Big[\sum_{i=1}^n \log^2 \lambda_i \Big]^{1/2}, \tag{4.64}$$

where λ_i, i=1,...,n are the real eigenvalues of $\mathbf{P}_1^{-1}\mathbf{P}_2$.

For each point $\mathbf{P} \in P(n)$, the set of tangent vectors at $\mathbf{P}$ defines a tangent space. Each tangent vector $\mathbf{S}_i$ can be seen as the derivative at $t = 0$ of the geodesic $\mathbf{\Gamma}_i(t)$ between $\mathbf{P}$ and the exponential mapping $\mathbf{P}_i = \text{Exp}_\mathbf{P}(\mathbf{S}_i)$, given by

$$\text{Exp}_\mathbf{p}(\mathbf{S}_i) = \mathbf{P}_i = \mathbf{P}^{\frac{1}{2}} \exp\left(\mathbf{P}^{-\frac{1}{2}} \mathbf{S}_i \mathbf{P}^{-\frac{1}{2}}\right) \mathbf{P}^{\frac{1}{2}}. \tag{4.65}$$

The inverse mapping is given by the logarithmic mapping defined as

$$\text{Log}_\mathbf{p}(\mathbf{P}_i) = \mathbf{S}_i = \mathbf{P}^{\frac{1}{2}} \log\left(\mathbf{P}^{-\frac{1}{2}} \mathbf{S}_i \mathbf{P}^{-\frac{1}{2}}\right) \mathbf{P}^{\frac{1}{2}}. \tag{4.66}$$
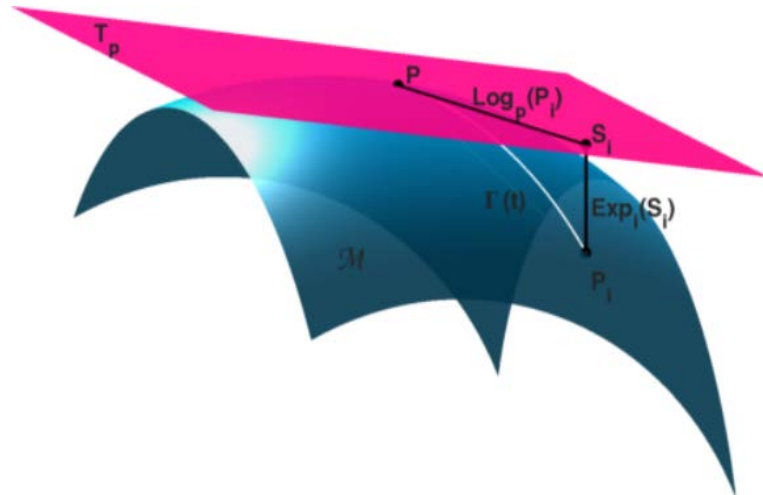
Fig. 4.12: Tangent space at point $\mathbf{P}$, $\mathbf{S}_i$ a tangent vector at $\mathbf{P}$ and $\mathbf{\Gamma}(t)$, the geodesic between $\mathbf{P}$ and $\mathbf{P}_i$ defined in the Riemannian manifold.

By using the Riemannian geodesic distance, the Riemannian mean of $I \geq 1$ SPD matrices is

given by

$$\mathfrak{G}(\mathbf{P}_1, ..., \mathbf{P}_I) = \operatorname*{argmin}_{\mathbf{P} \in P(n)} \sum_{i=1}^{I} \delta_R^2(\mathbf{P}, \mathbf{P}_i) \tag{4.67}$$

This mean is also referred to as the geometric mean. For a manifold of non-positive sectional curvature like $P(n)$, such local minimum exists and is unique [106]. However, there is no closed-form expression to compute the mean and optimization algorithms must be applied. An efficient iterative algorithm to compute the Riemannian mean of I SPD matrices is given in [107].

4.6.2.1.1 Classification in the Riemannian manifold

Algorithm 2 (MDRM) summarizes the process to obtain the Riemannian means and the corresponding distance between them and an unknown element, which are subsequently used to assign the corresponding class label. For a K-class problem, K Riemannian means are computed (one mean per class), and the label of an unknown element is assigned according to the closest Riemannian distance between itself and the computed means.

Algorithm 2 Minimum Distance to Riemannian Mean (MDRM)

Input: A set of $\mathbf{X}_i$ training trials of K different known classes.
Input: A testing data set comprising EEG trials of unknown class.
Input: $\mathcal{J}^{(k)}$, the set of indexes for trials belonging to the kth condition.
Output: $\hat{k}$, the estimated class of the trials comprised in the testing set.
    1) Compute the covariance matrices of the training trials $\mathbf{X}_i$ to obtain $\mathbf{P}_i$.
    2) Compute the covariance matrices of the testing trials to obtain $\mathbf{P}$.
    3) **for** $k = 1$ to K **do**
        $\mathbf{P}_{\mathfrak{G}}^{(k)} = \mathfrak{G}(\mathbf{P}_i, i \in \mathcal{J}^{(k)})$, use equation 4.67.
    end for
    4) $\hat{k} = argmin_k \delta_R(\mathbf{P}, \mathbf{P}_{\mathfrak{G}}^{(k)})$, use equation 4.64.
    5) **return** $\hat{k}$.

Figure 4.13 shows the features extracted by applying the MDRM algorithm over all data recorded from subject 3 of the 4-*class*, and by considering only motor imageries of the left and the right hand. It can be observed that separation of classes is well achieved, though there are several outliers that difficult the task. In particular, it is possible to observe for both classes a cluster of data that is shifted with respect to the rest. These elements correspond to first of the four runs that were conducted during the recording session. Presumably, there were changes in the electrodes impedance as the session advanced, and/or in the subject mental state.

4.6.2.2 CSP and Riemannian geometry

As mentioned, the MDRM algorithm uses the band-pass filtered EEG trials covariance matrices as descriptors, which are subsequently considered within the Riemannian manifold, and from
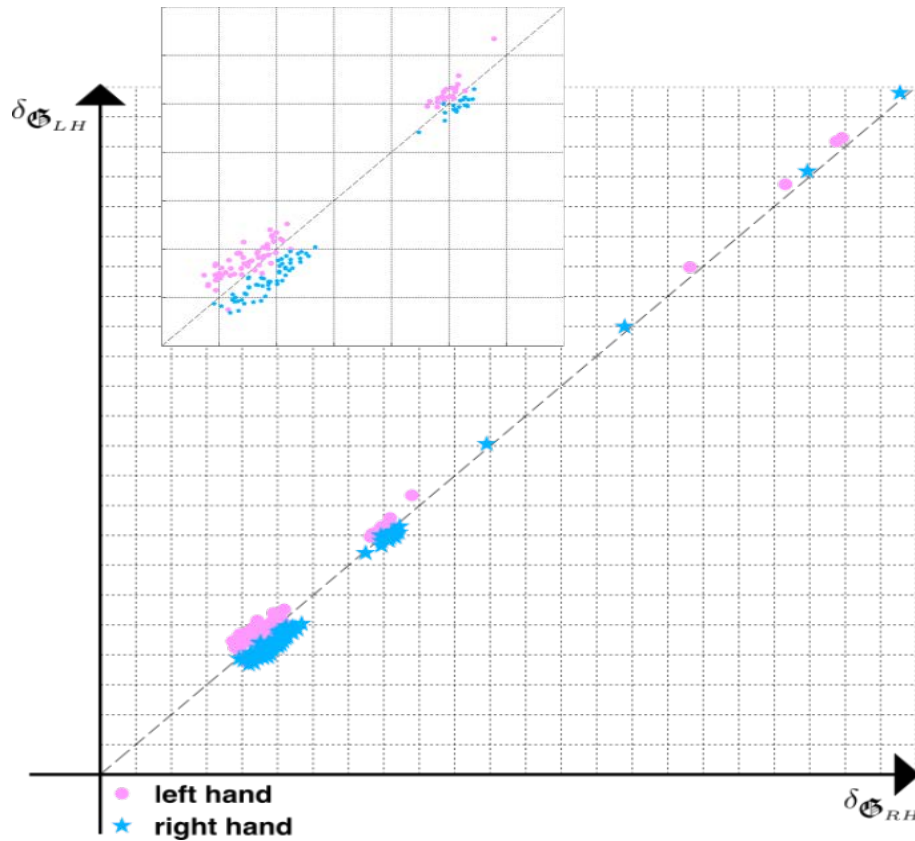
Fig. 4.13: MDRM classification. During the training phase, Riemannian means are computed for the left hand and right hand conditions. Subsequently, during validation, class labels are assigned according to the shortest Riemannian distance separating the input covariance matrices from the aforementioned class means. Note that this example is just for illustration purposes and that the same data were used to train/validate the classification model. Data belong to subject 3 from the 4-*class* database. Signals were band-pass filtered in the frequency range from [8-30 Hz].

which it is possible to extract the spatial information without using spatial filtering. Nevertheless, the spatial filters generated over two classes with the CSP algorithm extract information that is relative to both conditions, since they attempt to enhance uncorrelated information. Since the covariance matrices of the CSP projected trials preserve the differences attained through the spatial filtering, we have considered that it is advantageous to apply the CSP algorithm before mapping them into the Riemannian space in order to generate more discriminative features. In other words, the spatial information that is obtained by only considering the band-pass filtered EEG covariance matrices within the Riemannian space, as it is done by the MDRM method, offers independent information of each class. Whereas the CSP projected covariance matrices provide uncorrelated information from both classes.

Figure 4.14 shows the same data that were used in the example illustrated in Figure 4.13 of the previous section after applying the MDRM algorithm over the CSP projected trials covariance matrices. It can be observed that separation between classes has been considerably enhanced.
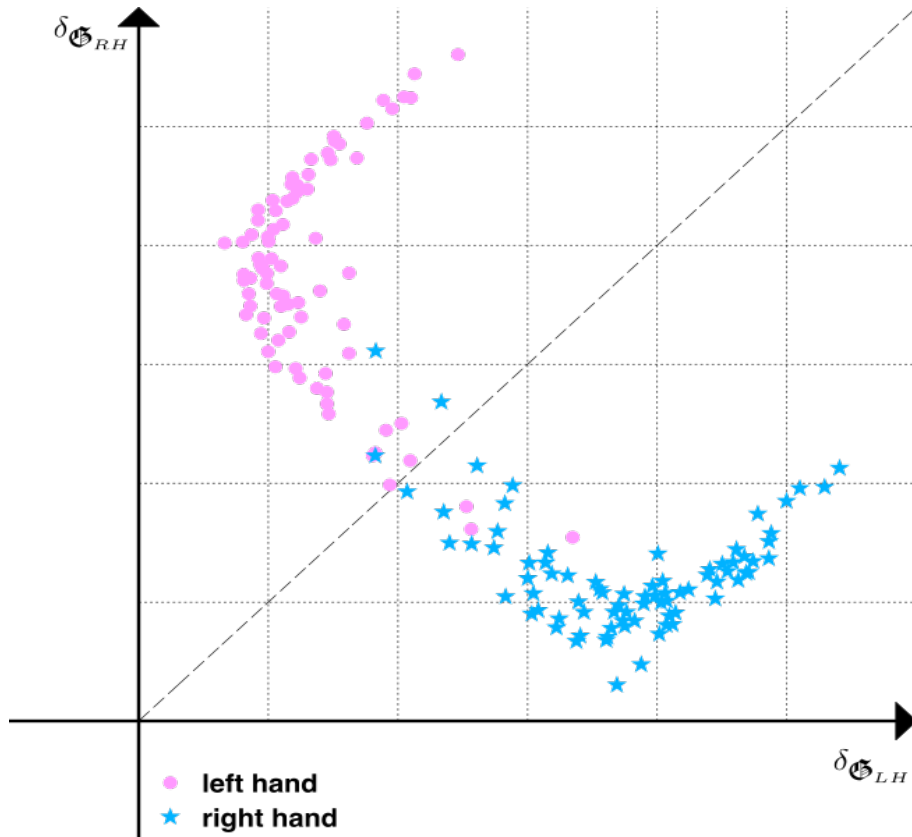
Fig. 4.14: MDRM classification in combination with the CSP algorithm. During the training phase, Riemannian means are computed over the CSP projected trials for the left hand and right hand conditions. Subsequently, during validation, class labels are assigned according to the shortest Riemannian distance separating the input covariance matrices computed over the CSP projected trials from the aforementioned class means. In contrast to the standard MDRM algorithm, distances are more separated. Note that this example is just for illustration purposes and that the same data were used to train/validate the classification model. Data belong to subject 3 from the 4-*class* database. Signals were band-pass filtered in the frequency range from [8-30 Hz].

<div style="text-align: right; font-size: 3em; color: gray;">**5**</div>

Multiclass and multilabel approaches

Contents

Along this chapter we will introduce the grouping methods for classification that have been developed in order to solve the multilabel classification problem that arises from the consideration of the combined motor imageries included in the 8-*class* database, in fact, two of these methods are also pertinent for the motor tasks comprised in the 4-*class* database. To this end, before going further, it is important to remember that, on the one hand, it is necessary to extract the features presenting the most discriminative information for contrasting two or more conditions, so that redundancies are eliminated and specific patterns enhanced. And on the other hand, once such features have been extracted, the predicted labels can be inferred by applying a classification algorithm that separates the features according to their position within the classification space (see sections 4.1 and 4.6 for further details).

In the first section we will introduce the concept of multiclass approach in the context of classification by describing three classical methods, namely, one-versus-one, one-versus-all, and

hierarchical algorithms. These methods will be subsequently used as benchmarks to compare the performance of the algorithms contributed by the present work. Such contributions will be presented in the second section as multilabel approaches for sensorimotor rhythms discrimination. The basis of each one of these new methods lies in the results that were presented along chapters 2 and 3, where we have suggested that EEG data recorded during a combined motor imagery can be modelled as the superposition of the activity that is independently generated by the sources associated to the body parts engaged in the task. Upon this generalization we have proposed a convenient manner to carry out data partition, so that the entire problem is reduced into a series of binary classification tasks. This strategy has been applied in three new different methods that vary in the manner of how the features extracted from each binary problem are classified, namely, One-step Multilabel (OsM) approach, Hierarchical Multilabel (HM) approach, and One-step Hiearchical Multilabel (OsHM) approach.

An important aspect of this way of grouping classes is that it optimizes the use of the CSP method described in section 4.1.3, which as mentioned consists of a convenient approach to differentiate sensorimotor rhythms that has the drawback of being suitable for separating only two classes. In this regard, in contrast to the one-versus-one and one-versus-all approaches, the proposed methods generate features throughout the same training process for all classes, so that the distances separating elements from the decision boundaries within the classification space present an appropriate scale that favors their subsequent classification.

5.1 Multiclass approaches

As mentioned in section 4.6, the problem of classifying an element in order to assign the corresponding class label consists in determining on which side of the decision surface it is located. For 2-class problems this is a relative simple task that can be solved by assessing the sign returned by the evaluation of the function that defines the decision boundary over the input element. Nevertheless, when the classification task involves more than 2 classes, this problem results more complicated. In fact, in this case it is necessary to generate multiple decision surfaces for separating all the considered classes, and to assign the corresponding labels according to the relative position of the input vectors to all boundaries within the classification space.

There are many possible ways for separating classes in order to set down the decision boundaries. Each approach varies in the number of functions that, according to the amount of classes, are required to divide the classification space into the associated regions. In this regard, the best approach is the one that satisfactorily adapts itself to the distribution of data while demanding a reasonable number of computations. In the following sections we will describe three of the most common methods, namely, one-versus-one, one-versus-all, and hierarchical approaches.

5.1.1   One-versus-one approach

As the name suggests, the one-versus-one approach consists in generating a model to independently separate each one of the groups. To this end, it is necessary to train $K(K-1)/2$ different models, where K represents the number of classes, on all possible pairs of groups and then, in the case of classification, to assign the labels of the testing set according to the class having the highest number of "votes" [87]. This technique has the drawback of leading to ambiguities in the resulting classification. However, this issue does not represent its biggest inconvenience; since it can be tackled by generating convex regions, as described in section 4.6.1.1. In fact, the main problem is that, for a large K, such solution requires a considerable amount of time for setting all the required models. Similarly, for the evaluation of testing data, significantly computation is required.

Figure 5.1 shows a representation of the one-versus-one approach for a classification problem involving 4 different classes. To solve such a problem $K(K-1)/2 = 6$ models are required.
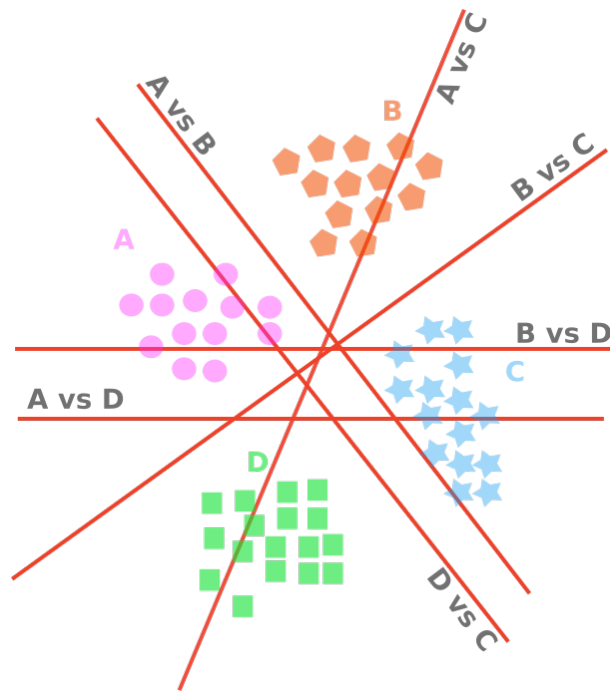
Fig. 5.1: One-versus-one classification for a 4-class problem. $K(K-1)/2$ different models, where K represents the number of classes, are generated to separate all possible pairs of groups. Finally, labels of the validation set are assigned according to the class having the highest number of "votes".

5.1.2   One-versus-all approach

Another common approach consists in constructing K separate models, where K represents the number of classes, in which the kth model is generated by grouping within one group all elements belonging to class C_k and, in the second class, all elements from the remaining $K-1$ classes. This is known as the one-versus-all approach. However, using the decisions of the individual classifiers

can generate inconsistent results; in which an input is assigned to multiple classes simultaneously. This problem is sometimes addressed by choosing the label associated to the largest distance separating the input data from the decision boundary among all models [108]. Unfortunately, this heuristic approach presents the inconvenience that the different models were trained on different tasks, and there is no guarantee that the distances generated by all models will have appropriate scales. Another problem that has the one-versus-all approach is that the training sets are unbalanced. For instance, if we have four classes each with equal numbers of training data elements, then the individual classifiers are trained on data sets comprising 75% of samples belonging to the group containing the $K-1$ classes, and only 25% samples of the group involving only the C_k class, which affects the original symmetry of the task.

Figure 5.2 shows a representation of the one-versus-all approach for a classification problem involving 4 different classes. To solve such a problem only $K = 4$ models are required.
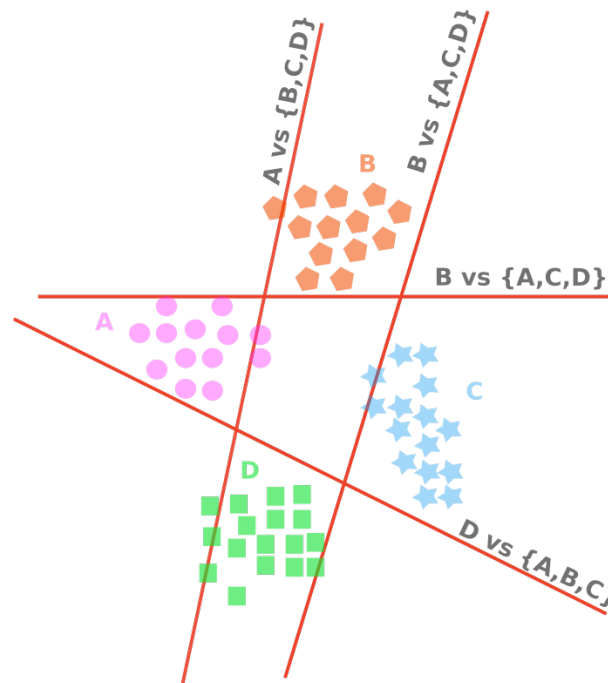
Fig. 5.2: One-versus-all classification for a 4-class problem. Consists in constructing K separate models, where K represents the number of classes, in which the kth model is generated by grouping within one group all elements belonging to class C_k and, in the second class, all elements from the remaining $K-1$ classes.

5.1.3 Hierarchical approach

Hierarchical methods allow solving multiclass problems by using a tree of binary classifiers, whose root discriminates between two groups; each one containing a half of the classes. Each succeeding node includes again only one half of the classes from the selected group, and the process is recursively repeated until each node contains a single class, from which the final decision can be

inferred [109].

The hierarchical approach can be a very convenient method, since even when several models are required to be generated during the training stage, only a few of them are used during the validation stage. However, there are two important drawbacks to be considered; first, if there is a single erroneous prediction throughout the classification chain, the final result will be wrong. And second, given that from the second stage all models are selected according to the previous decision, all the subsequent stages have to be applied consecutively, which can lead to long computations.

Figure 5.3 shows a representation of the hierarchical approach for a classification problem involving 4 different classes. As it can be observed, two stages are required; the first one consists of a model that selects 2 of the classes, which are subsequently evaluated by another model that is applied in the second and last stage in order to make the final decision.
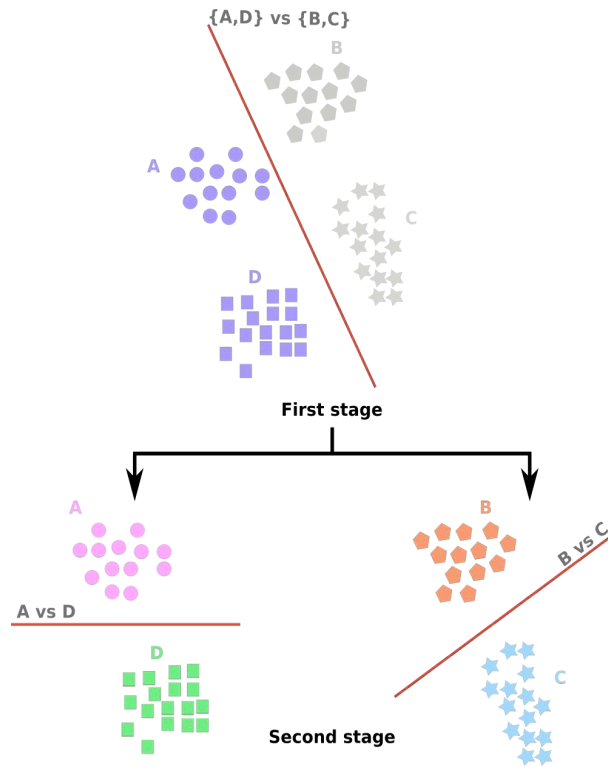
Fig. 5.3: Hierarchical approach for a 4-class problem. The first stage of the classification is shown on top, where one group corresponds to the ensemble of classes A and D, and the other group to the ensemble of classes B and C. The second stage of the process is shown on the bottom; during which, according to the predicted label assigned in the previous stage, one of two models is selected for inferring the final decision.

5.2 Multilabel approaches

Conventional single-label classification involves learning from a set of examples that are associated with a single label l from a set of disjoint labels L, $|L| > 1$. As mentioned, if $|L| = 2$, the learning

task is called binary classification, while if $|L| > 2$, then it is called multiclass classification. In multilabel problems the examples are associated with a set of labels $Y \subseteq L$ [110]. In the past, multilabel classification was mainly motivated by medical diagnosis tasks. For example, it can be the case that a patient suffers from both hypertension and atrial fibrillation at the same time.

Nowadays multilabel classification methods are increasingly required by modern applications, such as protein function classification, and among others, music categorization [111]. Essentially, there are two ways for addressing multilabel classification; *problem transformation methods*, which as its name indicates consists in transforming the multilabel task either into one or more single-label classifications problems, and *algorithm adaptation methods*; which extend specific learning algorithms in order to handle multilabel data directly.

In the case of motor imagery-based BCI systems, the multilabel approach arises when combined movements are included in the paradigm; which as mentioned, allows to drastically increase the number of afforded commands. Now, if we consider the assumption that was justified by the analysis presented in sections 2.6.3 and 3.1.3 for the 4-*class* and 8-*class* databases respectively, it turns out that there is a particular way for grouping classes that allows to define a problem transformation method based on the manner of how the activity is distributed within the different sources. In simple terms, this activity appears as an ERD modulation over the regions whose associated limb is engaged in the motor imagery. In the case of combined imaginations, such modulation can be characterized as the superposition of the activity that is independently generated by each source during simple tasks [21], [22]. With this in mind we have proposed to group EEG data separately for each activity source; so that, all motor imageries involving the use of the associated body part, are gathered together into one class (hereinafter to be referred to as C_{ERD}), and the remaining conditions that do not include it in another class (hereinafter to be referred to as C_{IDLE}). In this way, it is possible to reduce the entire problem into a series of binary tasks; which simplifies the complexity of the classification and boost feature extraction by allowing the use of the CSP algorithm, which is only suitable to discriminate between two classes. By using this type of partition we have implemented three new different methods that vary in the manner of how the features extracted from each binary problem are classified, namely, One-step Multilabel (OsM) approach, Hierarchical Multilabel (HM) approach, and One-step Hierarchical Multilabel (OsHM) approach.

Before introducing the aforementioned problems we present, in order to simplify their description, a convention for labeling each one of the different tasks according to the presence/absence of motor imagery-related activity at each one of the main sources. If the sources are active, a value of "1" is used to represent them; otherwise, the value corresponds to "0". The number assigned to each source is used to generate a P-bit word, with P representing the number of body parts considered by the paradigm (2 for the 4-*class* database, and 3 for the 8-*class* one), where the less significant bit corresponds to the area around C4 (right side) and the most significant bit to the area around C3 (left side). Hereafter, in order to describe the proposed methods, we will refer to each class by considering the decimal equivalent of the aforementioned convention (see figure

5.4 for a complete association between motor tasks and labels).

Fig. 5.4: Labeling convention. The proposed convention assigns a value of "1" to represent the sources that are active in terms of motor imagery-related activity; otherwise, the value corresponds to "0". The number assigned independently to each source is used to generate a 3-bit word, whose decimal equivalence will be used when referring to each one of the classes considered in the paradigm. Note that the same convention is applied to the 4-*class* database, for which the sources associated to the left hand and right hand are used to generate a 2-bit word, where the less significant bit is associated to the use of the left hand, and the most significant bit to the use of the right hand.

5.2.1 One-step Multilabel (OsM) approach

The One-step Multilabel (OsM) method considers, in the case of the 4-class problem formulated by the 4-*class* database, two feature extraction modules to independently determine whether the sources linked to the left hand and right hand are engaged during a motor task (see Figure 5.5). Similarly, in the case of the 8-class problem given by the 8-*class* database, besides these two models there is a third one to analyse the activity generated by the source associated to the feet (see Figure 5.6). Each module is defined as a binary problem where one class corresponds to the group of motor tasks showing ERD modulation at the corresponding source (i.e., motor imageries involving the use of the limb associated to that source), and the other class to the group of motor tasks that do not generate such activity and present similar patterns to the idle state (i.e., motor imageries that do not involve the use of the limb associated to that source).

The whole procedure involves two feature extraction models in the case of the 4-class problem (compared to the six and four models that are required to solve the same problem by using the one-versus-one and one-versus-all approaches respectively), and three feature extraction models for the 8-class one (compared to the 28 and 8 models that are required to solve the same problem by using the one-versus-one and one-versus-all approaches respectively). In both cases, outputs are concatenated to form a single vector, so that only one classification model is required.

When applied in combination with the CSP method, and by selecting 2 pairs of filters (4 in total), for each one of the two binary problems devoted to solve the 4-class task, the OsM approach generates a total of 8 features per trial. Whereas, in the case of the 8-class problem, the resulting vectors comprise 12 features per trial.
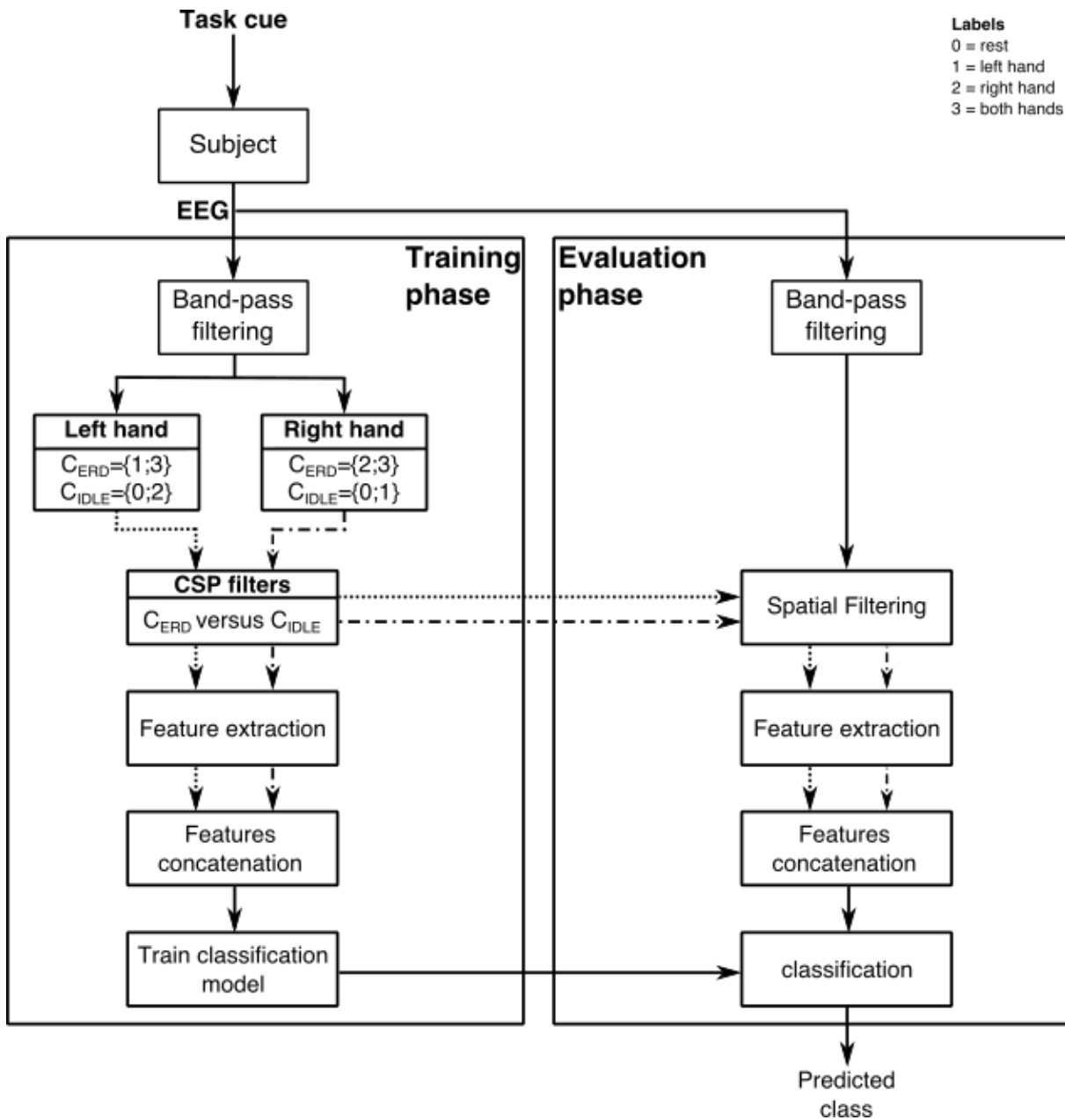
Fig. 5.5: Architecture of the One-step Multilabel Approach (OsM) for solving the 4-class problem. Two different binary problems are generated around each activity source, where one class corresponds to all motor imageries involving the use of the associated limb (denoted by C_{ERD}), and the other class to those motor imageries that do not involve it (denoted by C_{IDLE}). The resulting features are concatenated and used to train/validate a single classification model.

5.2.2 Hierarchical Multilabel (HM) approach

The second method consists of a hierarchical decision process that comprises three feature extraction models distributed in two consecutive stages for solving the 4-class problem given by the 4-*class* database (see Figure 5.7), and seven feature extraction models placed within three
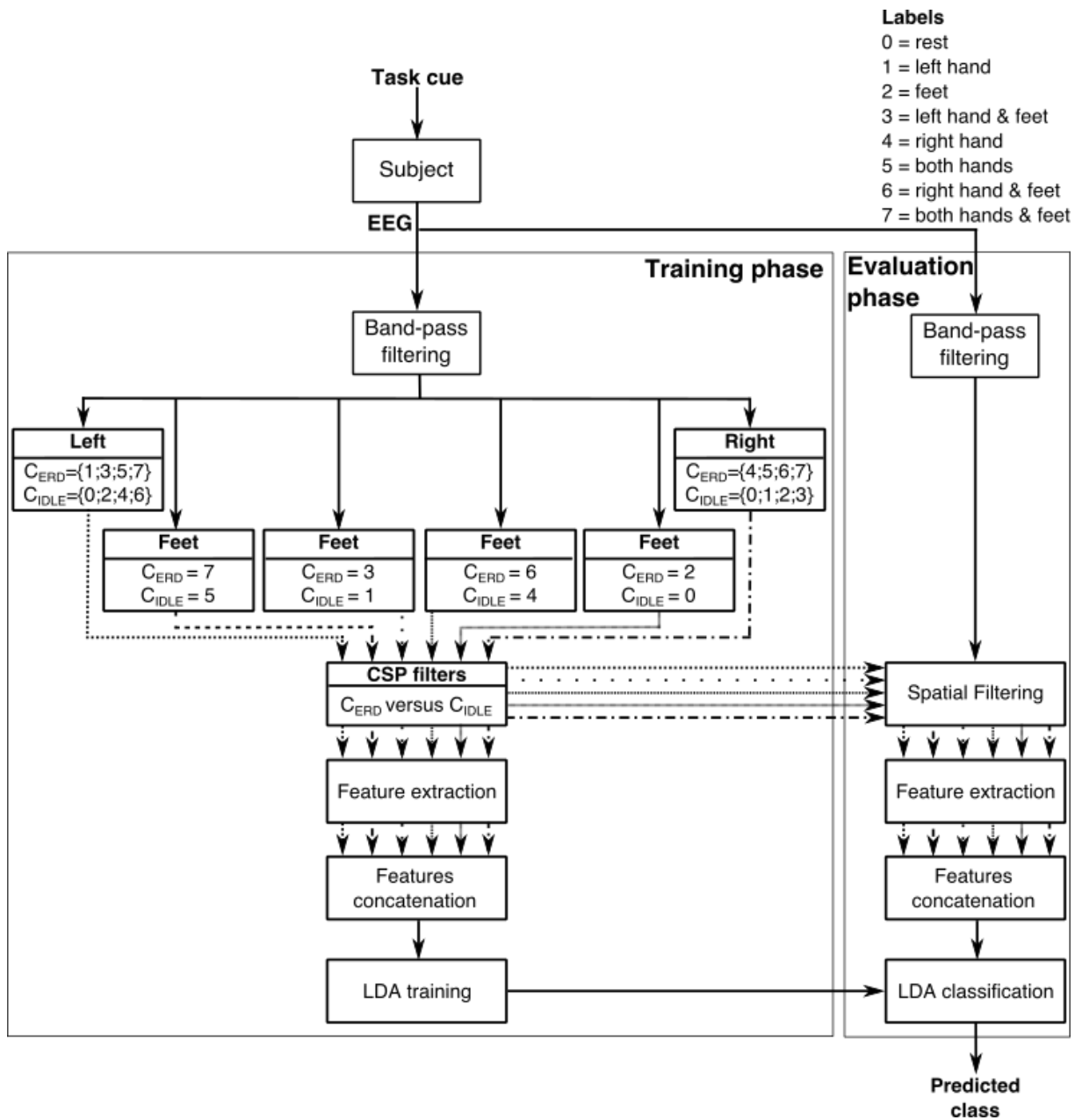
Fig. 5.6: Architecture of the One-step Multilabel Approach (OsM) for solving the 8-class problem. Three different binary problems are generated around each activity source, where one class corresponds to all motor imageries involving the use of the associated limb (denoted by C_{ERD}), and the other class to those motor imageries that do not involve it (denoted by C_{IDLE}). The resulting features are concatenated and used to train/validate a single classification model.

consecutive stages for solving the 8-class problem formulated by the 8-*class* database (see Figure 5.8). Such stages constitute a tree of binary classifiers whose root discriminates between two groups containing each one a half of the classes. Each succeeding node includes only the half of the classes that was selected in the previous stage, and the process is recursively repeated until the last node contains only two classes, from which the final decision can be inferred. Data

partition is carried out in a similar manner as the OsM method; so that the presence/absence of motor imagery-related activity associated to one of the hands is analyzed during the first stage in order to determine whether or not it is engaged in the motor task. To this end, two groups are conformed by gathering together the motor tasks involving the selected hand in the ERD-class, and by putting together the motor tasks that do no involve it in the IDLE-class. Note that the hand selection to start the classification process is arbitrary, and that it is possible to start either with the left or the right hand. Once the decision is inferred, half of the classes are discarded. After this decision there are only two classes left in the case of the 4-class problem, whereas for the 8-class task there are four classes remaining (i.e., in both cases either the group of motor tasks involving the selected hand or the group that does not include it). Based on this criterion, two feature extraction and classification models are generated in the second stage; from which, according to the previous prediction, only one of them is applied during the validation phase. In this part of the process the aim is to identify whether or not the source linked to the hand that has not been analyzed presents ERD activity, given that the hand selected after the first stage has been already recognized as active or inactive in terms of motor imagery-related activity. In the case of the 4-class problem the second stage corresponds with the last one, given that each one of the two models corresponds to a binary problem including only two classes, from where the final decision can be inferred. One of these models correspond to the one formulated by the engagement of the hand analyzed in the previous stage, in which case the goal is to determine whether or not the other hand is active, so that the predicted label would result in the use of both hands if the second hand is also found to be engaged, or only the use of the hand that was already detected in the previous step if the second hand was found to be inactive. In the same way, the second model is generated under the assumption that the hand analyzed during the first stage was labelled as inactive, so that the goal is to determine whether the other hand is also inactive, so that the prediction would correspond to the rest state, or whether the second hand is active, in which case the predicted label would be assigned according to the manner in which data partition was proposed in the formulation of the problem. In the case of the 8-class problem, the second stage corresponds to an intermediate process that aims at determining the state of the second hand before to start analyzing the activity generated by the feet. Thus, if the inference of the first step has considered the ERD-class, the corresponding model generates a new ERD-class by gathering together the two motor tasks that involve the use of both hands; and in a new IDLE-class the two motor tasks that involve the use of the hand selected in the previous model but that do not involve the use of the hand under the current analysis. Contrarily, if the inference of the previous step has considered the IDLE-class, the corresponding model creates a new ERD-class by putting together the two motor tasks that involve the use of the hand under analysis but that do not include the use of the hand selected by the previous step; and a new IDLE-class by gathering together the motor tasks that do not involve the use of any hand. After the second stage there are only 2 classes remaining, from which the final decision can be inferred in the third and last step. To this end, there are four different instances that are determined by

the combination of decisions made by the preceding stages, each one of them defining a binary problem that considers the activity generated by the source linked to the feet. Thus, in the case that any hand was found to generate ERD activity, the classification task consist of distinguishing between rest and feet. Contrarily, if only the left hand was found to be involved in the motor task, the classification aims at discriminating between left hand and left hand in combination with feet. In the same way, if only the right hand has generated ERD activity, the task consists of classifying right hand versus right hand in combination with feet and, finally, if both hands were found to be involved during the motor task, the goal is to distinguish between both hands and both hands in combination with feet.

5.2.3   One-step Hierarchichal Multilabel (OsHM) approach

The third method is a combination of the first two approaches; it implements the same binary problems related to the sources associated to the hands, as it was done with the OsM method, and incorporates the four instances associated with the source linked to the feet that were formulated by the HM algorithm. The whole procedure involves 6 feature extraction models (compared to the 28 and 8 models that are required to solve the same problem by using the one-versus-one and one-versus-all approaches respectively), whose outputs are concatenated to form a single vector, so that only one classification model is required. In this way, by considering all the possibilities involving the use of the feet, more discriminative features are generated in comparison with those obtained by the OsM approach. And, given that all features are considered together, it is not necessary to apply different classification stages as it was done with the HM approach, which speeds up the process and helps to compensate outliers.

When applied in combination with the CSP method, and by selecting 2 pairs of filters (4 in total) for each one of the 6 binary problems, the OsHM approach generates a total of 24 features per trial. Figure 5.9 shows the architecture of the OsHM approach for solving the 8-class problem formulated by the 8-*class* database. Note that in this case there is no adaptation for the 4-*class* database since, given that it only includes the use of the hands, such approach is equivalent to the OsM method.
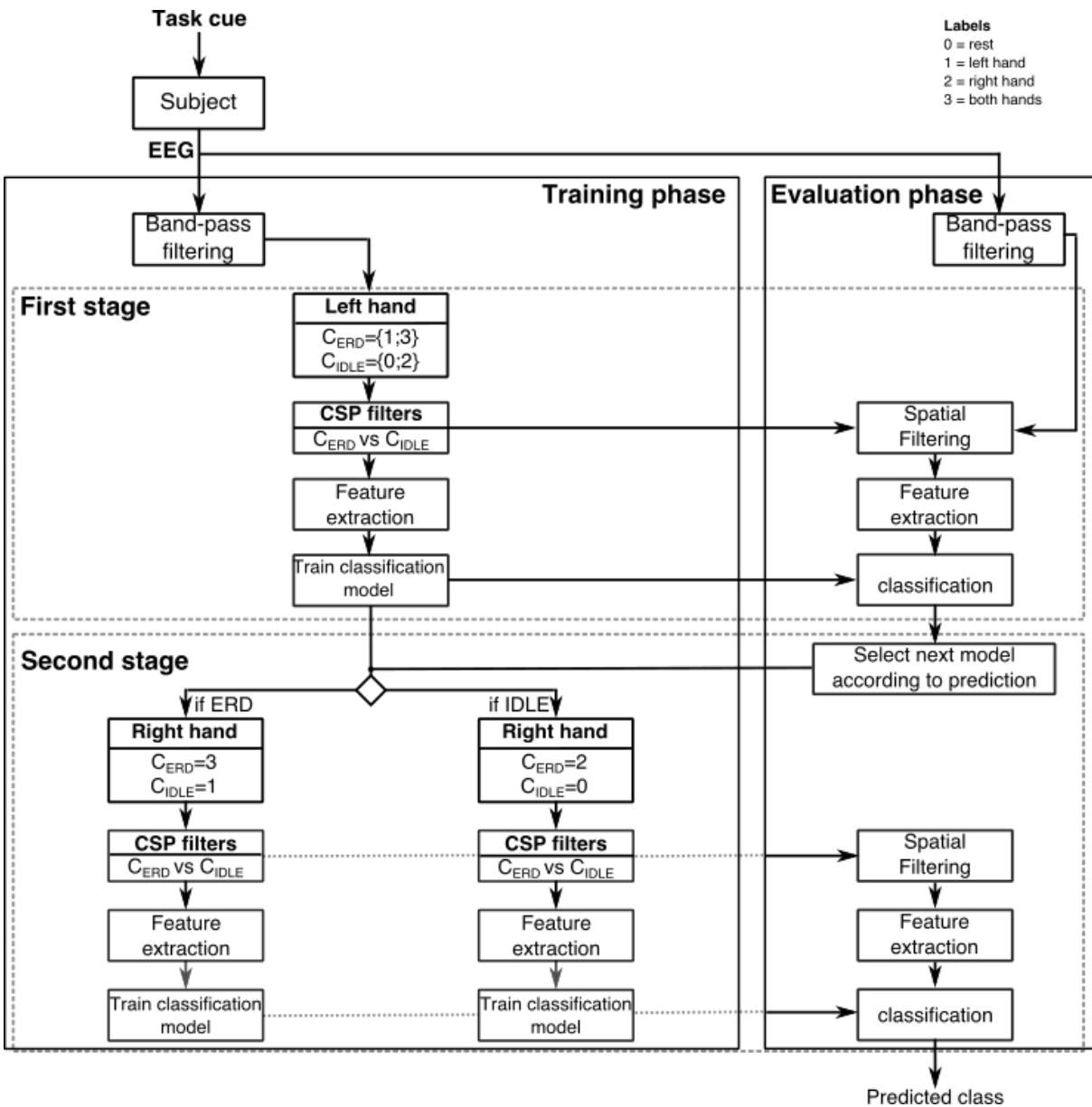
Fig. 5.7: Architecture of the Multilabel Hierarchical approach (MH) for solving the 4-class problem. During the training phase three feature extraction and classification models are implemented; one model for the first stage, which is trained to discriminate between the two motor tasks involving the use of the selected hand (denoted by C_{ERD}) from the two motor tasks that do not involve it (denoted by C_{IDLE}), and two binary models in the second stage which are trained to discriminate between the motor tasks involving the hand that has not been analyzed (denoted by C_{ERD}) from those that do not involve it (denoted by C_{IDLE}), given that the hand selected for the first stage has been already recognized as active or inactive in terms of motor imagery-related activity. During the evaluation phase only two models are needed, which are selected according to the prediction of the previous stages. Note that this scheme illustrates the case where data partition is carried out by analysing first the engagement of the left hand.
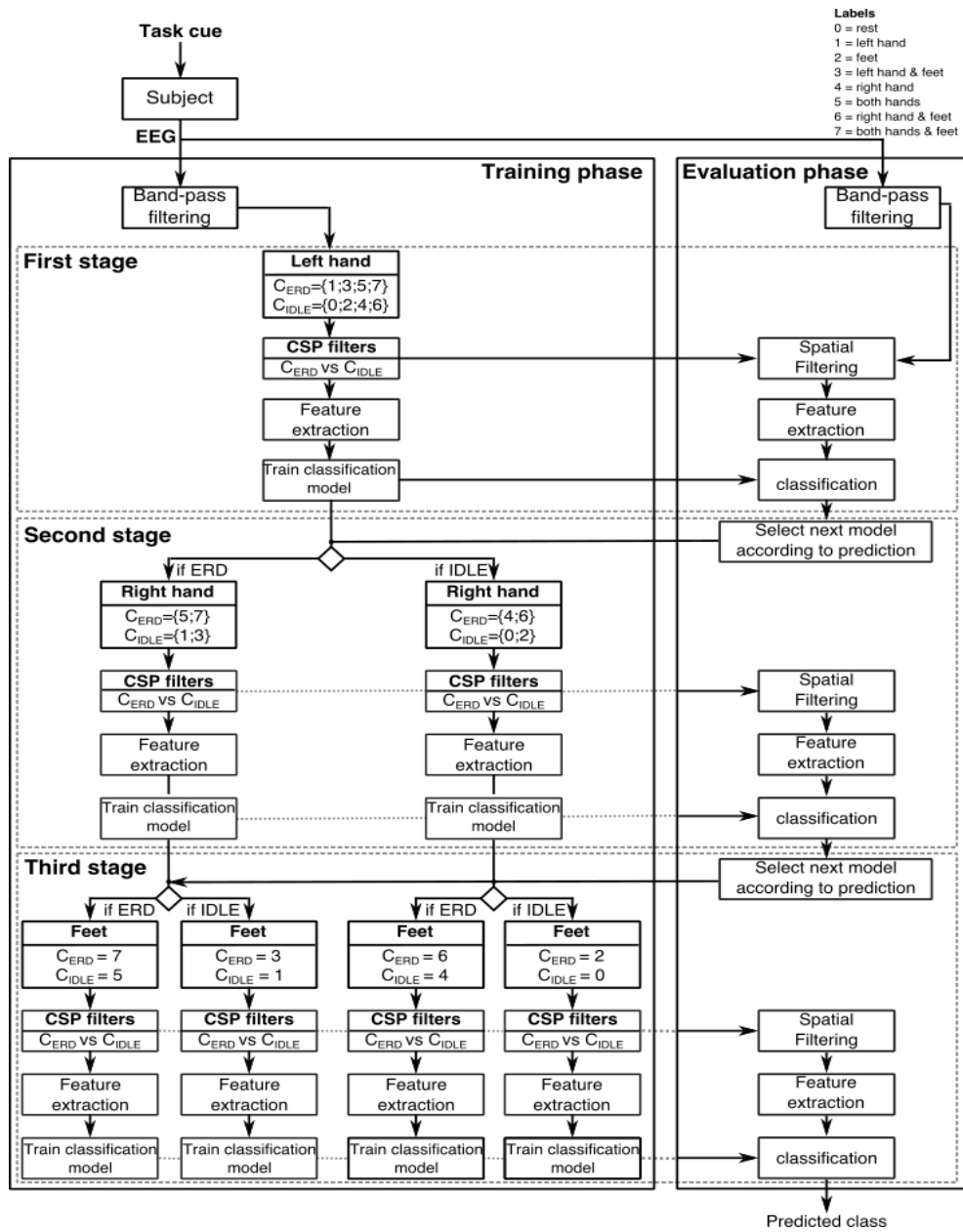
Fig. 5.8: Architecture of the Multilabel Hierarchical approach (MH) for solving the 8-class problem. During the training phase seven feature extraction and classification models are implemented; one model for the first stage, which is trained to discriminate between the four motor tasks involving the use of the selected hand (denoted by C_{ERD}) from the four motor tasks that do not involve it (denoted by C_{IDLE}), two binary models in the second stage which are trained to discriminate between the motor tasks involving the hand that has not been analyzed (C_{ERD}) from those that do not involve it (C_{IDLE}), given that the hand selected for the first stage has been already recognized as active or inactive in terms of motor imagery-related activity, and finally, four binary models in the third stage to determine whether or not the feet are involved in the motor task, given that the hands have been already identified as active or inactive. During the evaluation phase only three models are needed, which are selected according to the prediction of the previous stages. Note that this scheme illustrates the case where data partition is carried out by analysing first the engagement of the left hand.
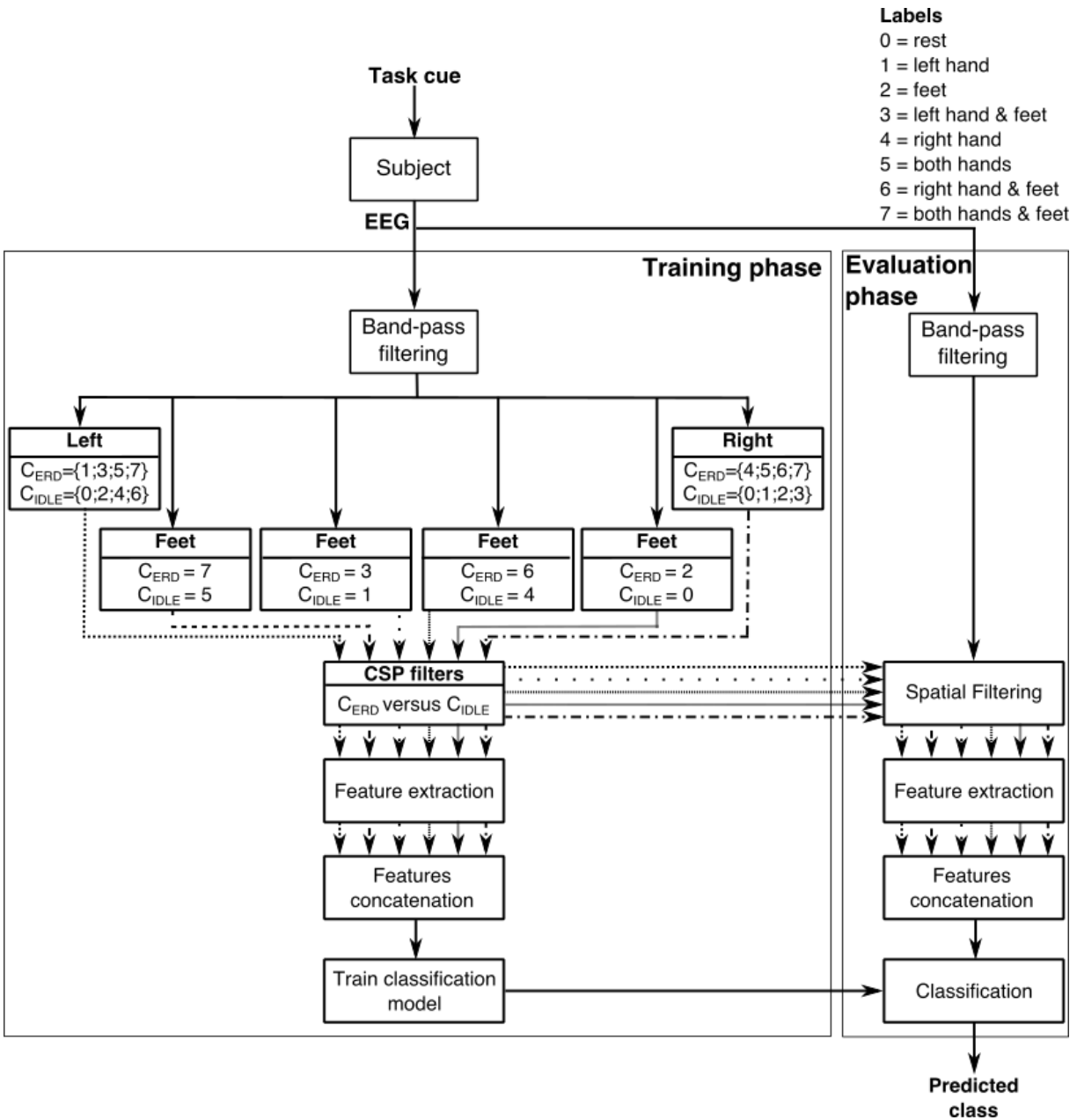
Fig. 5.9: Architecture of the One-step Multilabel Approach (OsM) for formulated for the 8-*class* database. Three sets of 4 CSP filters are computed from the band-pass filtered EEG signals of the training dataset. Each one of them is obtained by solving a binary problem at each one of the three main activity sources associated to the body parts considered by the paradigm. All MIs inducing ERD modulation are gathered together in one class denoted as C_{ERD} and, all MIs that do not show ERD activity and present similar patterns to the idle state are grouped in the so-called C_{ERS}. The 12 resulting features for each trial are concatenated and used to train an LDA classifier to solve the 8-class problem involving all motor tasks. During the evaluation phase, the unknown elements are band-pass filtered and projected into the three feature spaces to subsequently concatenate the resulting features and classify them according to the LDA model.

6

Experimental results

Contents

Along this chapter we will present the offline results generated after applying the feature extraction methods discussed in chapter 5 on the 4-*class* and 8-*class* databases. For classification, Linear Discriminant Analysis and Support Vector Machines were applied on the 4-*class* database in order to evaluate the performance of the proposed methods by exploring different solutions (excepting the Riemannian distance-based schema, which do not require a complementary classifier). Results obtained by the Support Vector Machines algorithm showed a considerably lower performance than the one generated by the Linear Discriminant Analysis method, presumably because the corresponding models were over-trained. Thus, for the 8-*class* database only results obtained by LDA are presented.

To start with, we specify the parameters that were used for signal preprocessing (i.e., filtering parameters and time window length), as well as the electrode montage that has been applied during

the recording sessions. Furthermore, we also enumerate the characteristics of the classification functions, as well as the complexity parameters that were estimated for the optimization of the different kernel functions considered for the SVM algorithm applied on the 4-*class* database, namely, linear, polynomial, and radial basis function. Subsequently, we describe the cross-validation procedures that were implemented on both databases for assessing all methods performance as well as for estimating the aforementioned complexity SVM parameters.

After indicating all experimental specifications we present, in terms of the mean accuracy among subjects, the offline results that were generated from applying each one of the feature extraction methods described along chapters 4 and 5, the detailed results obtained for each subjects, are presented in appendix C.

6.1 Experimental parameters

In order to carry out an EEG motor imagery-based session, it is necessary to set a series of parameters that depend on the characteristics of the experiment. For instance, it is necessary to choose the amount and position of the recording electrodes; which as discussed in section 2.2, should be located in such a way that the primary motor cortex is covered, so that the elicited activity generated by the considered activity sources is well conducted. Moreover, if the paradigm is asynchronously implemented, it is required to establish the time window in relation with the task cue for classification, which should include the period where the signal patterns associated to the motor task are prevalent (i.e., not before it starts reflecting the associated activity, nor after it begins to stabilize). Additionally, the frequency range for the band-pass filtering must be selected in a way that all the components containing the patterns of interest are considered, whereas those related to the background activity are suppressed.

The tuning of these parameters represents a research concern, since they have direct influence on the system performance and, given that there is a huge inter-subject variability with respect to the characteristics of the brain signals [94], many efforts have been made in order to develop automated algorithms for their optimization [90], [112]. However, these aspects are out of the scope of the present work and we have decided to use the same parameters along all subjects and for both databases, which have been selected according to the observations described in chapters 2 and 3. All the selected values for the mentioned parameters that were used during the experimental setup are listed below

— **Recording electrodes**
  — Fp1; Fp$_z$; Fp2; F$_z$; FC5; FC3; FC1; FC$_z$; FC2; FC4; FC6; C5; C3; C1; C$_z$; C2; C4; C6; CP5; CP3;CP1; CP$_z$; CP2; CP4; CP6, and P$_z$, re-referenced with respect to the common average reference across all channels and located over the international 10-10 system positions to cover the primary sensorimotor cortex (see fig. 2.7)
— **Filtering**

— Each trial was independently filtered to better emulate an online setup by using a 5th order Butterworth filter within the frequency range from 8 to 30 Hz.

— **Time window size**

— A 4-second window starting 0.5 s after the task cue was considered for classification.

6.1.1 Classification algorithms

As mentioned in chapter 4, besides feature extraction, it is necessary to implement a classification algorithm that according to the position of the feature vectors situated within the classification space assigns their corresponding class label, so that input data are translated into the commands that allow controlling the interface. On this regard, we have decided to evaluate different classification algorithms to assess the performance of the contributed approaches presented in chapter 5 in comparison with classical methods. To this end, for the 4-*class* database,we present all results obtained by Linear Discriminant Analysis and Support Vectors Machine classification, both of which were described in sections 4.6.1.3 and 4.6.1.4, respectively. As mentioned, in the case of the 8-*class* database we restrict the evaluation only to LDA classification, since the SVM algorithm generates poorer performance.

For Linear Discriminant Analysis classification we consider two cases; the premise of equivalent covariance matrices of the involved classes (LDA), and the assumption of their specific covariance matrices (QDA). All feature extraction methods (excepting those using a Riemannian distance-based scheme) were evaluated by using the Linear Discriminant Analysis classification under a one-versus-one and one-versus-all approach (see sections 5.1.1 and 5.1.2, respectively).

On the other hand, for the Support Vectors Machine classification applied on the 4-*class* database, we consider three different kernel functions, namely, linear, polynomial, and radial basis function. Each one of these models requires a set of complexity parameters. For instance, it is necessary to estimate the box constraint, which controls the trade-off between under- and overfitting. For the linear kernel this is the only parameter that we have considered for the model optimization, whereas for the polynomial and radial basis functions, in addition to the box constraint, we address, respectively, the optimization of the polynomial order and the gamma parameter (i.e., the inverse of the radius of influence of samples selected by the model as support vectors). All parameters have been estimated by using the cross-validation procedure that is described in the following section.

For the 4-*class* database, all feature extraction methods (excepting the Riemannian distance-based methods, and those using the ACSP algorithm, since it generates complex features requiring a special strategy for applying the SVM approach, which is out of the scope of the present work), were evaluated in combination with the SVM algorithm under a one-versus-one and one-versus-all approach by using the three aforementioned kernel functions.

6.2 Cross-validation

Due to the overfitting problem, performance obtained on the training set is not a good indicator of predictive performance on unseen data. When the amount of data is sufficiently large, then one approach simply consists in using some of the available data to train a range of models, or a given model with a range of values for its complexity parameters, and then to compare them on independent data, sometimes called *validation set*, and select the one generating the best results [87]. In order to generate robust models, it is important to use as much of the available data as possible for training. However, in many applications the available amount of data is limited, and hence, if the validation set is small, it will produce a noisy estimate of predictive performance.

One possible solution for this inconvenience is to use cross-validation; which consists of using a proportion $(s-1)/s$ of the available data for training while making use of all of the remaining data to asses performance. When data is particularly limited, it may be appropriate to consider the case $S = N$, where N is the total number of data points, which gives the *leave-one-out* technique.

An important drawback of cross-validation is that the number of training runs that must be performed is increased by a factor of S, and this can be problematic for models in which the training is itself computationally expensive. A further problem is that there might exist multiple complexity parameters for a single model. In these cases it becomes necessary to apply a *nested* cross-validation, which consists of carrying out another cross-validation process over the $(s-1)/s$ training data to evaluate the impact of different values for the involved parameters. Exploring combinations of settings for such parameters could, in the worst case, require a number of training runs that is exponential in the number of parameters.

6.2.1 Cross-validation on the 4-class database

As mentioned in section 2.6.1, the 4-*class* database contains 4 runs of 20 trials per class for each subject, which totals 80 trials per class. Given that we have observed some intra-subject variability over the distribution of data belonging to different runs (see figure 4.13 as an example), which might be presumably due to mental state changes and/or fluctuations in the electrodes impedance throughout the recording session, we have decided to implement a 4-fold cross-validation using the trials of an entire run (20 trials per class) as the testing set, and the trials of the 3 remaining runs (60 trials per class) as the training set. Such procedure is repeated so that all runs are used to train and test the corresponding models. In this way, we can better emulate the variability that it is expected to be found across different sessions, and to generate results whose performance is independent from a (un)fortunate (un)favorable data partition.

Figure 6.1 shows the 4-fold cross validation process that has been implemented to evaluate the classification performance obtained by all the discussed feature extraction algorithms applied in combination with the LDA and QDA methods, as well as for the distance-based approaches. Note that none of these classifiers requires any complexity parameter, and therefore the cross-validation

process pursues only to evaluate their performance without needing a nested procedure. Once all possible choices have been implemented, the performance scores for the 4 folds are averaged to obtain the performance estimate for each model.

Fig. 6.1: 4-fold cross-validation process for assessing performance over the 4-*class* database for the discussed feature extraction methods in combination with the LDA and QDA algorithms, as well as for the distance-based approaches. Throughout each fold, the testing set (represented in dark blue) was constituted by all trials belonging to one of the runs (20 trials per class), and the training set (represented in dark blue) was integrated by the remaining 3 runs (60 trials per class). After all folds, all possible choices are implemented and the performance scores for the 4 folds are averaged.

In contrast to the linear discriminant analysis algorithm, classification based on SVM requires different complexity parameters that depend on the selected kernel function. In order to determine the optimal values, one possible solution involves performing a nested cross-validation; which consists in partitioning again the training set generated at each iteration to perform a new cross-validation process for assessing the model performance over a range of values. Figure 6.2 shows the nested procedure that has been implemented on the 4-*class* database to evaluate the performance of the discussed methods applied in combination with SVM classification. As in the previous case, we have decided to use a 4-fold cross-validation considering, at each iteration, all trials belonging to one run as the testing set, and the 3 remaining ones as the training set. In addition, towards determining the optimal complexity parameters, another cross-validation procedure is performed over the training set generated at each iteration, for which a range of different values is evaluated over the classification model in order to select the optimal ones.

As discussed in section 4.6.1.4, regardless the selected kernel function, the SVM algorithm requires the estimation of the box constraint, which controls the trade-off between under- and overfitting. For the linear kernel this is the only parameter that we have considered for the model optimization, whereas for the polynomial and radial basis functions, in addition to the box constraint, we address, respectively, the optimization of the polynomial order and the gamma parameter (i.e., the inverse of the radius of influence of samples selected by the model as support vectors). For both the box constraint and the gamma parameter we have evaluated a series of values covering the range from 10^{-2} to 10^2 [103], [113], whereas for the polynomial order we have restricted the analysis, in order to avoid overfitting, within a range from 2 to 5. As illustrated, the nested cross-validation involves also 4 iterations, for each one of them, the different values within the proposed range are evaluated. When more than one parameter need to be estimated, then the corresponding range is evaluated for each combination of values. Finally, the optimal parameters are selected from the highest performance score to implement, by using the original training set, the optimal classification model which, in turn, is evaluated on the original testing set.
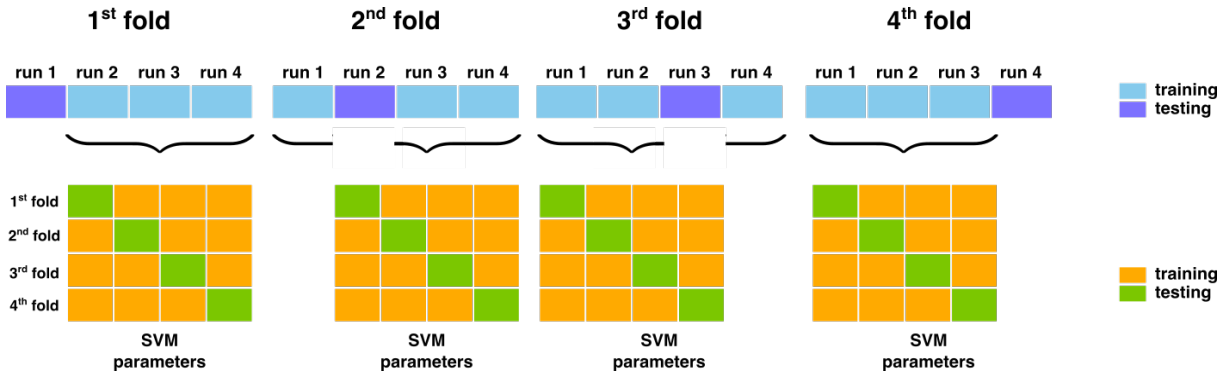
Fig. 6.2: Nested cross-validation process for assessing performance over the 4-*class* database for the discussed feature extraction methods in combination with the SVM algorithm. Performance is evaluated through a 4-fold cross-validation. Given that the SVM algorithm requires a series of complexity parameters, another 4-fold cross validation is performed over the training set generated at each iteration, for which a range of different values are evaluated over the classification model in order to select the optimal ones. When more than one parameter need to be estimated, then the corresponding range is evaluated for each combination of values. Once such values have been determined, the model is implemented by using the original training set and evaluated on the original testing set.

6.2.2 Cross-validation on the 8-class database

In contrast to the 4-*class* database, the 8-*class* one contains rather less trials per class due to the increase in time that it takes to accomplish such a high number of different motor tasks; which causes users fatigue and their loss of concentration. As mentioned in section 3.1, the 8-*class* database contains 3 runs of 10 trials per class, which sums 30 trials per class for a total of 240 trials recorded in the whole session for each subject. Because of this limited amount of data, we have decided to apply a different cross-validation approach that the one described in the previous section, which consists of a 30-fold procedure that considers for each iteration one trial per class (8 trials in total) as the testing set, and the remaining 29 trials per class (232 trials in total) as the training set. This approach allows using more trials for training than the ones that would be available by following the same scheme proposed for the 4-*class* database, where the trials belonging to one run were considered as the testing set, and the remaining ones as the training set. In fact, we could have implemented the leave-one-out technique discussed at the beginning of this section to generate more robust models, however, since it drastically increases the computational expenses, we have decided to use the mentioned procedure illustrated in figure 6.3. As it can be observed, throughout each fold the testing set was constituted by one trial consecutively selected from each class (8 trials in total), and the remaining 29 trials per class (232 trials in total) as the training set. After all folds, all possible choices are implemented and the performance scores for the 30 folds are averaged.
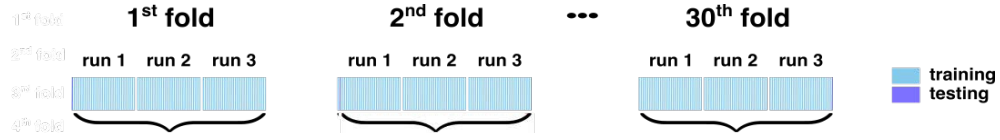
Fig. 6.3: 30-fold cross-validation process for assessing performance over the 8-*class* database for all discussed methods in combination with the LDA and QDA algorithms, as well as for the distance-based approaches. Throughout each fold, the testing set (represented in dark blue) was constituted by one trial consecutively selected from each class (8 trials in total), and the remaining 29 trials per class (232 trials in total) as the training set. After all folds, all possible choices are implemented and the performance scores for the 30 folds are averaged.

6.3 Results: summary

Finally, in the following section, we summarize all the reported results on tables 6.1 and 6.2 for the 4-*class* and the 8-*class* databases respectively, where performance is shown in terms of the grand average accuracy across subjects. Note that in cases where some of the classification functions were not applied, either because the features were complex or because the feature method corresponded to a distance-based algorithm, this has been indicated as "NA" (non applicable). In the same way, we have used the terms "OVO" and "OVA" for referring to the one-versus-one and one-versus-all approaches respectively.

Table 6.1: Classification results on 4-class database

				Accuracy		
				SVM		
Method	**LDA**	**QDA**	**Linear**	**Polynomial**	**RBF**	**MD**
CSP$_{OVO}$	53.5±9.3%	51.7±8.9%	50.6±8.1%	45.7±9.9%	47.2±8.9%	NA
CSP$_{OVA}$	53.9±9.5%	52.4±8.7%	41.8±4.7%	39.1±6.2%	30.8±1.8	NA
ACSP$_{OVO}$	53.7±7.9%	52.1±9.2%	NA	NA	NA	NA
ACSP$_{OVA}$	54.3±8.5%	53.7±9.7%	NA	NA	NA	NA
FBCSP$_{OVO}$	51.6±8.8%	49.3±7.9%	48.2±7.4%	48.0±7.3%	41.1±7.5%	NA
FBCSP$_{OVA}$	52.2±7.3%	51.5±7.9%	50.2±5.1%	47.8±5.7%	38.2±9.2%	NA
FBACSP$_{OVO}$	50.1±8.4%	50.5±8.5%	NA	NA	NA	NA
FBACSP$_{OVA}$	48.9±9.2%	48.0±8.3%	NA	NA	NA	NA
CSPbyJAD	49.7±11.1%	45.4±11.0%	44.8±9.9%	40.9±8.5%	41.6±7.8%	NA
MDRM	NA	NA	NA	NA	NA	45.4±6.5
CSP$_{OVO}$/MDRM	NA	NA	NA%	NA	NA	48.00±8.50%
CSP$_{OVA}$/MDRM	NA	NA	NA	NA	NA	53.25±8.19%
OsM$_{CSP}$	53.6±8.6%	48.7±8.7%	50.1±8.2%	46.7±8.4%	50.7±8.9%	NA
OsM$_{ACSP}$	54.0±9.3%	51.1±10.3%	NA	NA	NA	NA
OsM$_{FBCSP}$	52.7±7.4%	48.9±7.7%	46.6±6.1%	34.2±6.5%	48.4±7.0%	NA
OsM$_{FBACSP}$	51.6±8.8%	49.3±7.9%	NA	NA	NA	NA
OsM$_{MDRM}$	NA	NA	NA	NA	NA	44.7±6.4%
OsM$_{CSP/MDRM}$	NA	NA	NA	NA	NA	48.0±8.5%
HM$_{CSP}$/left	51.3±8.3%	49.8±7.7%	50.5±9.5%	49.5±9.6%	50.3±9.1%	NA
HM$_{ACSP}$/left	53.1±9.1%	52.6±9.5%	NA	NA	NA	NA
HM$_{FBCSP}$/left	51.4±6.2%	49.7±7.8%	35.7±3.1%	36.5±4.5%	36.1±4.2%	NA
HM$_{FBACSP}$/left	47.1±8.9%	47.5±6.8%	NA	NA	NA	NA
HM$_{MDRM}$/left	NA	NA	NA	NA	NA	46.7±6.0%
HM$_{CSP/MDRM}$/left	NA	NA	NA	NA	NA	50.1±8.1%
HM$_{CSP}$/right	50.3±8.1%	50.1±8.5%	49.7±6.6%	49.5±7.8%	50.5±9.6%	NA
HM$_{ACSP}$/right	53.0±10.3%	51.8±8.8%	NA	NA	NA	NA
HM$_{FBCSP}$/right	**55.2±8.6%**	51.2±8.2%	36.4±4.1%	36.2±3.9%	36.9±4.3%	NA
HM$_{FBACSP}$/right	49.1±8.7%	49.7±8.5%	NA	NA	NA	NA

Table 6.2: Classification results on 8-class database

Method	Accuracy		
	LDA	**QDA**	**MD**
CSP$_{OVO}$	46.0±16.0%	45.0±15.7%	NA
CSP$_{OVA}$	42.5±15.5%	41.6±15.9%	NA
ACSP$_{OVO}$	47.0±16.4%	41.1±15.9%	NA
ACSP$_{OVA}$	43.6±16.7%	43.9±15.1%	NA
FBCSP$_{OVO}$	47.1±16.0%	45.8±16.1%	NA
FBCSP$_{OVA}$	40.4±16.3%	38.7±15.9%	NA
FBACSP$_{OVO}$	44.2±15.3%	37.6±15.5%	NA
FBACSP$_{OVA}$	36.8±15.2%	37.1±14.2%	NA
CSPbyJAD	44.9±14.0%	38.9±16.6%	NA
MDRM	NA	NA	40.3±13.7%
CSP$_{OVO}$/MDRM	NA	NA	41.5±15.1%%
CSP$_{OVA}$/MDRM	NA	NA	41.7±14.8%
OsM$_{CSP}$	47.7±16.2%	41.1±16.48%	NA
OsM$_{ACSP}$	**49.3±16.5%**	43.4±15.8%	NA
OsM$_{FBCSP}$	41.0±15.5%	37.4±15.1%	NA
OsM$_{FBACSP}$	39.8±14.9%	23.8±13.8%	NA
OsM$_{MDRM}$	NA	NA	34.1±12.9%
OsM$_{CSP/MDRM}$	NA	NA	37.5±12.9%%
HM$_{CSP}$/left	37.9±15.7%	36.8±14.7%	NA
HM$_{ACSP}$/left	43.2±16.4%	40.8±15.5%	NA
HM$_{FBCSP}$/left	42.5±16.1%	42.8±15.7%	NA
HM$_{FBACSP}$/left	42.3±15.9%	39.4±16.1%	NA
HM$_{MDRM}$/left	NA	NA	37.7±14.3%
HM$_{CSP/MDRM}$/left	NA	NA	41.5±15.3%
HM$_{CSP}$/right	43.5±16.6%	41.9±15.0%	NA
HM$_{ACSP}$/right	45.0±16.2%	41.1±16.0%	NA
HM$_{FBCSP}$/right	40.3±15.9%	37.8±15.4%	NA
HM$_{FBACSP}$/right	40.3±15.3%	37.8±15.4%	NA
HM$_{MDRM}$/right	NA	NA	35.1±14.6%
HM$_{CSP/MDRM}$/right	NA	NA	39.0±13.8%
OsHM$_{CSP}$	47.1±15.7%	43.83±17.3%	NA
OsHM$_{ACSP}$	46.7±16.3%	41.9±16.8%	NA
OsHM$_{FBCSP}$	44.1±14.4%	–	NA
OsHM$_{FBACSP}$	42.3±15.9%	–	NA
OsHM$_{MDRM}$	NA	NA	36.8±13.7%
OsHM$_{CSP/MDRM}$	NA	NA	37.0±15.1%

by the OsM, HM, and OsHM approaches in comparison to the classical one-versus-one and one-versus-all algorithms for each subject. In all cases features were extracted by using the CSP method in combination with LDA classification. In addition, Figure 6.5 shows the mean accuracy with the corresponding standard error of the mean (SEM) among all subjects for each one of the methods. And, in order to observe the performance of each method over the best subjects, Figure 6.6 present the mean accuracy together with the SEM among subject, 3, 10 and 11. Note that the

difference between methods is significant, and that the OsM approach considerably outperforms the classical ones.

Similarly, Figure 6.7 shows the same comparison by using the analytical version of the CSP algorithm and LDA classification. Figure 6.8 shows the mean accuracy with the corresponding standard error of the mean (SEM) among all subjects for each one of the methods. And, in order to observe the performance of each method over the best subjects, Figure 6.9 present the mean accuracy together with the SEM among subject, 3, 10 and 11. Note that the difference between methods is significant, and that the OsM approach considerably outperforms the classical ones.
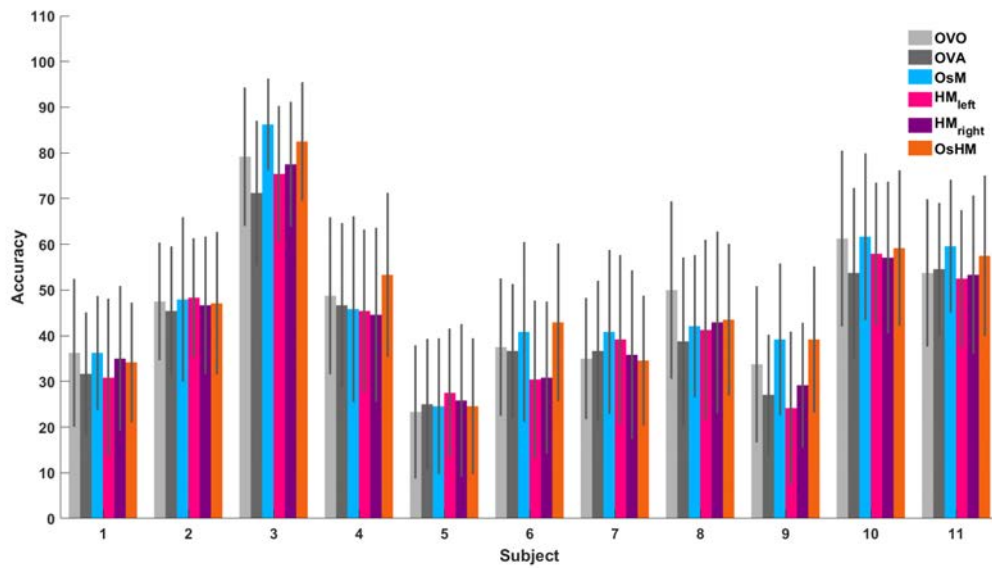
Fig. 6.4: Comparison of the accuracy results obtained by using the proposed methods and the one-versus-one and one-versus-all approaches in combination with the CSP algorithm for all subjects of the 8-*class* database. Note that classification was carried out by Linear Discriminant Analysis. Error bars indicate the Standard Deviation.
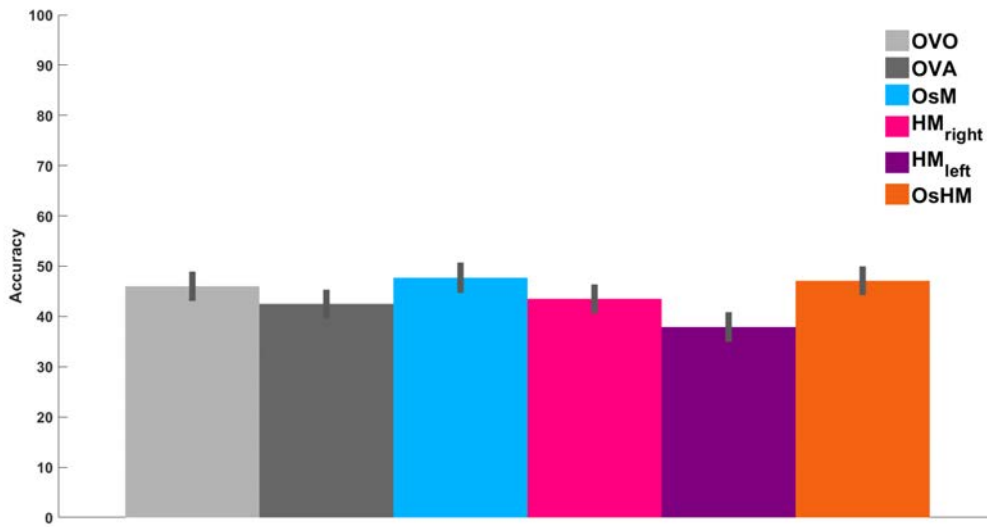
Fig. 6.5: Comparison of the mean accuracy results obtained by using the proposed methods in combination with the CSP algorithm for all subjects of the 8-*class* database. Error bars indicate the Standard Error of the Mean.
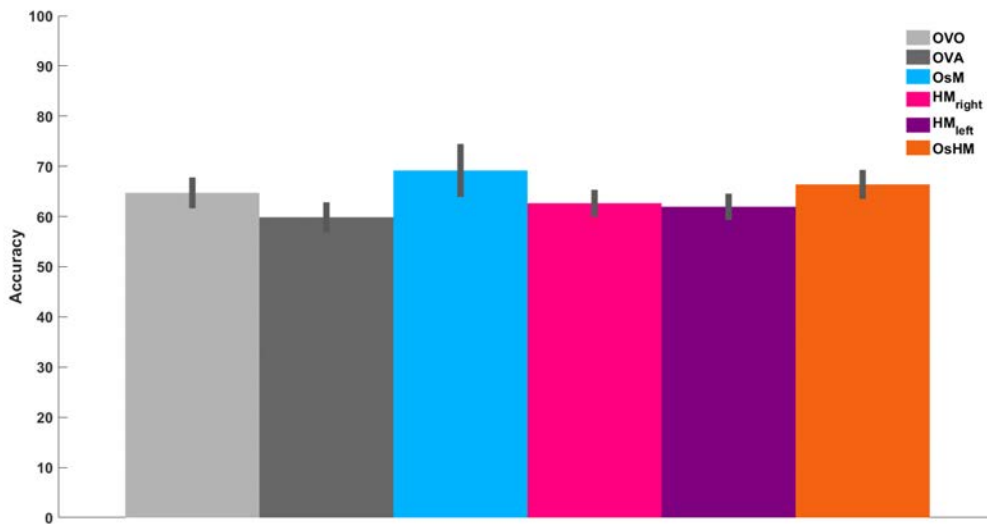
Fig. 6.6: Comparison of the mean accuracy results obtained by using the proposed methods in combination with the ACSP algorithm for all subjects of the 8-*class* database. Error bars indicate the Standard Error of the Mean.
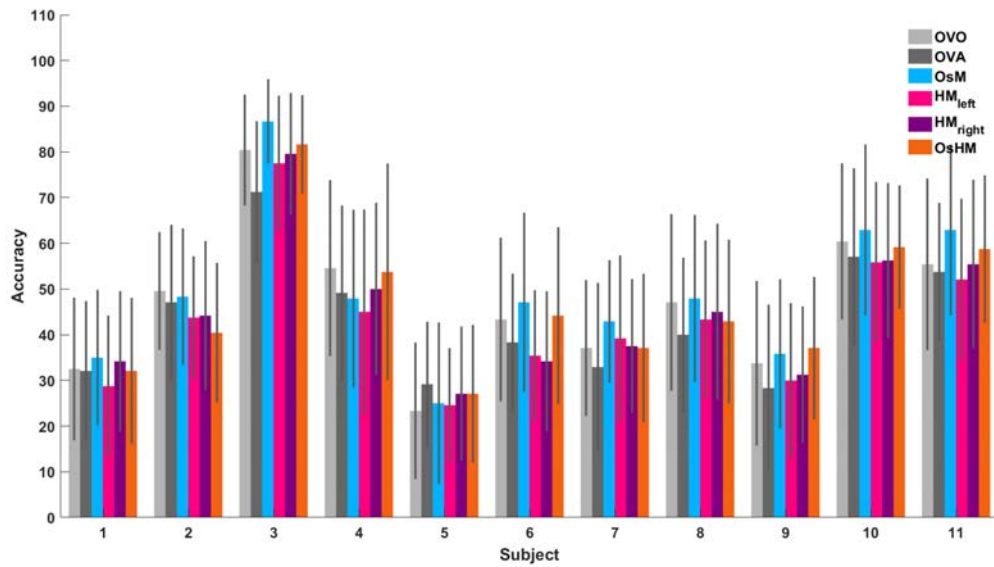
Fig. 6.7: Comparison of the accuracy results obtained by using the proposed methods and the one-versus-one and one-versus-all approaches in combination with the ACSP algorithm for all subjects of the 8-*class* database. Note that classification was carried out by Linear Discriminant Analysis. Error bars indicate the Standard Deviation.
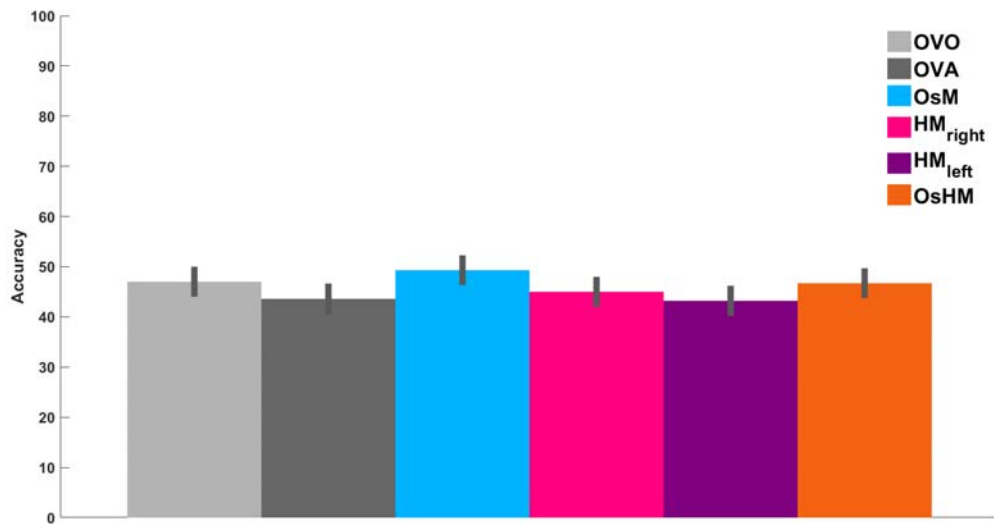
Fig. 6.8: Comparison of the mean accuracy results obtained by using the proposed methods in combination with the ACSP algorithm for all subjects of the 8-*class* database. Error bars indicate the Standard Error of the Mean.
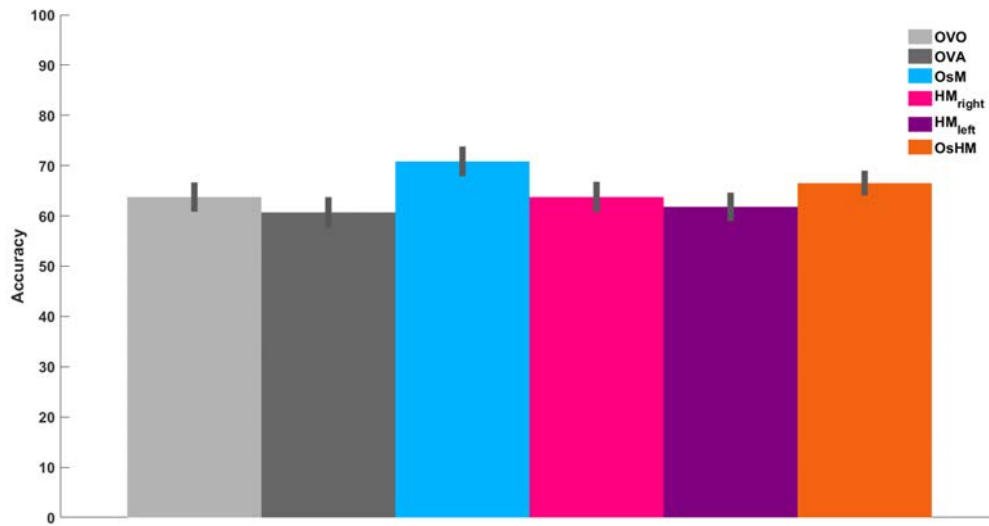
Fig. 6.9: Comparison of the mean accuracy results obtained by using the proposed methods in combination with the ACSP algorithm for all subjects of the 8-*class* database. Error bars indicate the Standard Error of the Mean.

Discussion

Throughout this section we will analyze the characteristics of the three contributed methods together with the results presented in the previous chapter. Since all classes contained in the 4-*class* database are also included in the 8-*class* one, the following analysis will be focused only on the latter. Moreover, given that the best results were obtained by applying all methods in combination with the CSP algorithm, we will center the discussion around this solution.

To start with, we will analyze the CSP patterns [114] and the CSP features generated by the OsM approach. Figure 10 shows a topographic representation of the first CSP patterns computed for each one of the three binary problems formulated by this method. Data were taken from subject 3 over the training set generated by one partition of the 30-fold cross validation described in section 6.2.2. It can be observed that there are strong weights over the motor cortex, more precisely over the sources that are associated to the body parts that are involved during the different motor tasks, as expected from the literature [5]; which confirms that such partition is neurophysiologically plausible.
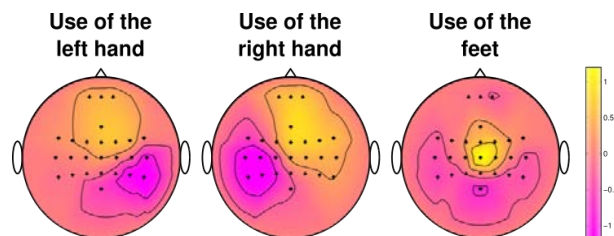
Fig. 10: Most discriminative CSP patterns obtained for subject 3 for each one of the binary problems formulated by the OsM approach. From left to right, the first CSP patterns generated for each binary problem considering the motor tasks inducing ERD activity from those that do not, at each one of the sources associated with the right hand (A); feet (B), and left hand (C). Note that the strongest weights are generated over the contralateral hemispheres.

For discussion purposes, Figure 11 presents the first two CSP features computed over the aforementioned training set by using the OsM approach. Note that these features correspond only to the first pair of filters and that some discriminative information may be missing. Also note that the proposed approach uses the concatenation of the three set of features by comprising them into a single vector. However, this representation is useful to observe that, in the case of the sources associated to the use of the left and right hands, features are well separated, on the contrary, features obtained from the problems involving the feet are overlapped for almost every

pair of conditions. This situation becomes clear when observing the analysis illustrated in figures 3.4 and 3.8, where we have seen that EEG activity around the central region is influenced by the surrounding sources. By considering this aspect, the last two methods have been conceived in such a way that the main decisions are based on the activity elicited by the sources related to the hands and, on a complementary level, by the activity that is linked to the feet.
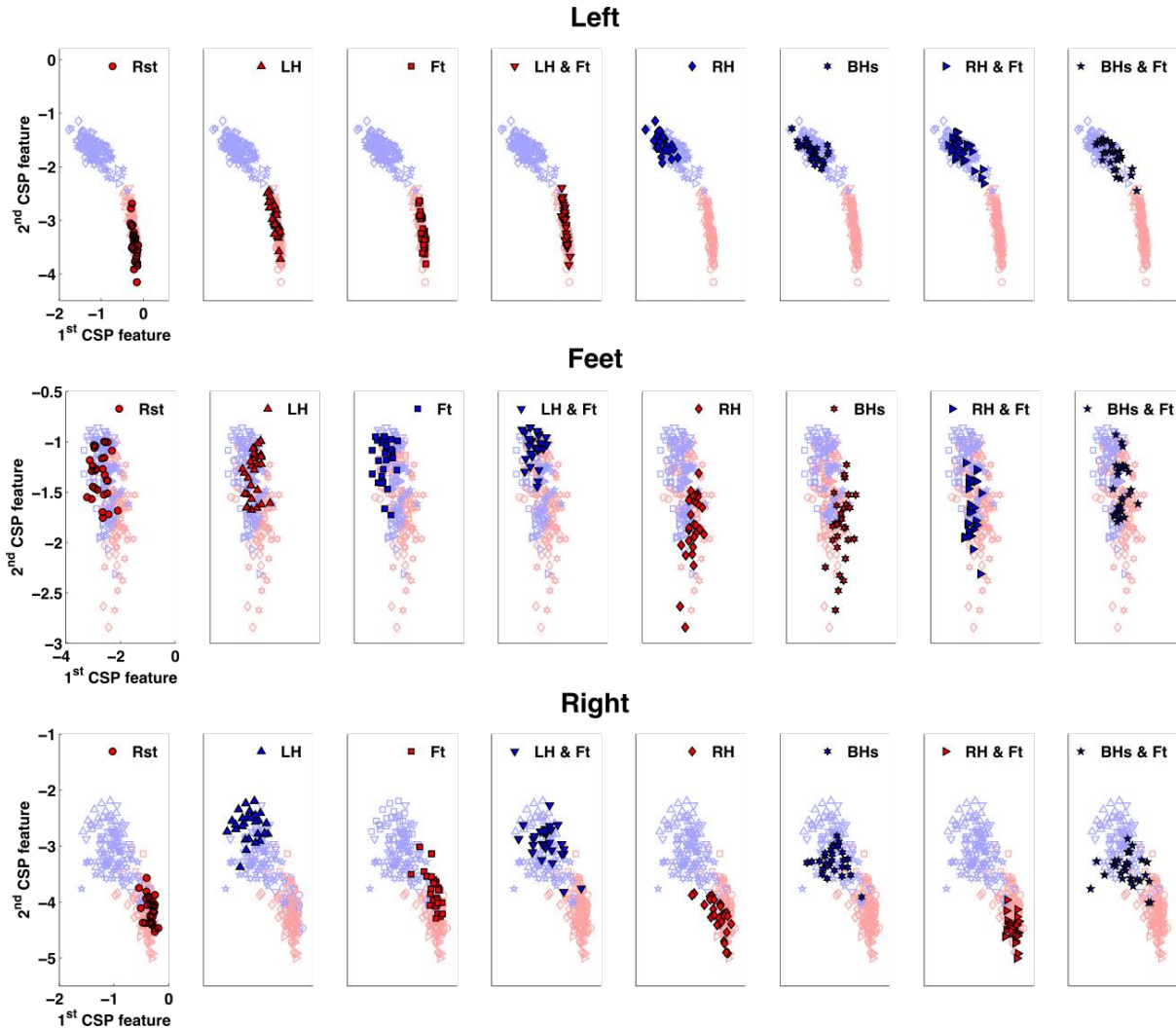
Fig. 11: First two features obtained with the CSP method for each one of the binary problems formulated by the OsM method for subject 3. Each row corresponds with the CSP features generated for each one of the 3 problems associated, from top to bottom, with the left hand, feet and right hand sources. From left to right the elements belonging to each one of the 8 different classes are highlighted in red if the corresponding source is not involved in the motor task, and in blue otherwise. Rst; LH; Ft; RH and BHs correspond, respectively, with rest; left hand; feet; right hand and both hands.

For the purpose of a fair comparison between the three methods, we will use the aforementioned data that were taken from the training set generated during the same partition of the 30-fold cross validation for subject 3. The first CSP patterns computed over these data for each one of

the 7 binary problems defined by the MH approach are shown, for the case in which the process starts by analysing the activity associated with the use of the left hand, in Figure 12 and, for the case in which the process starts by analysing the activity linked to the use of the right hand, are presented in Figure 13.
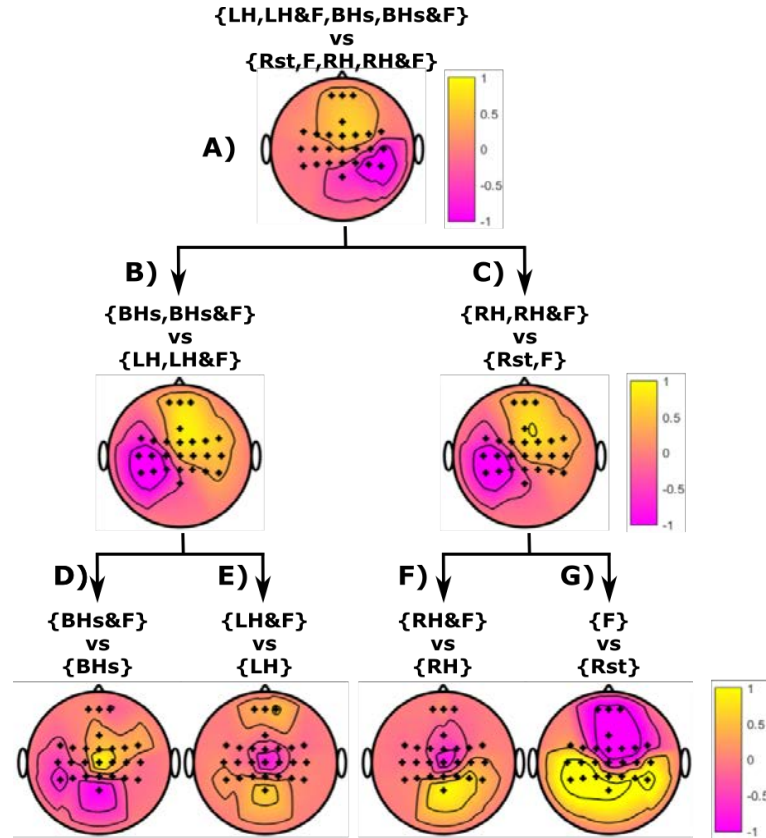
Fig. 12: Most discriminative CSP patterns obtained for subject 3 for each one of the binary problems formulated by the HM method starting with the left hand. From top to bottom the CSP patterns generated for the left hand (A); the right hand given that there was no ERD activity found in the source associated with the left hand (B); the right hand given that the source associated with the left hand was found to generate ERD activity (C), and the four instances formulated by the hierarchical approach involving feet: both hands versus both hands and feet (D); left hand versus left hand and feet (E); right hand versus right hand and feet (F), and rest versus feet (G).

As observed in the previous method, there are strong weights distributed around the associated activity sources over the motor cortex confirming that this partition is also plausible. Figure 14 shows, step by step, the first 2 features generated for each class across the 3 stages formulated by the HM approach. It can be observed that for all problems conditions are well separated and, contrarily to the first method, classification problems involving the feet show better discrimination; which is due to the fact that, for these cases, classes were conformed by using only 2 conditions, and hence more specific patterns are considered.
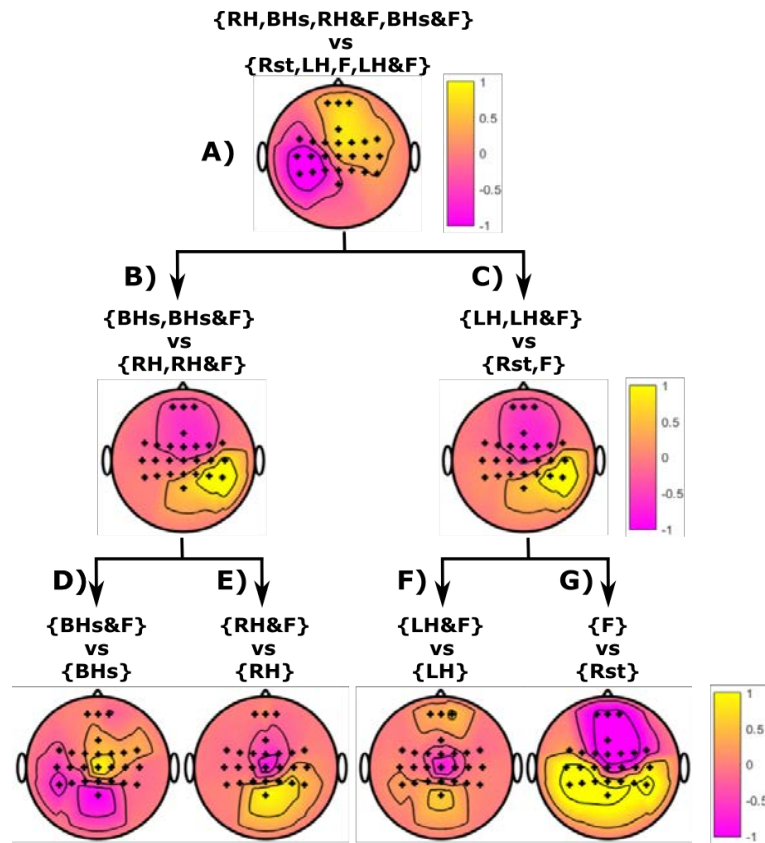
Fig. 13: Most discriminative CSP patterns obtained for subject 3 for each one of the binary problems formulated by the MH method starting with the right hand. From top to bottom the first patterns generated for the right hand (A); the left hand given that there was no ERD activity found in the source associated with the right hand (B); the left hand given that the source associated with the right hand was found to generate ERD activity (C), and the four instances formulated by the hierarchical approach involving feet: both hands versus both hands and feet (D); right hand versus right hand and feet (E); left hand versus left hand and feet (F), and rest versus feet (G).

Even though features are well separated after applying the HM approach, classification results do not outperform the OsM algorithm. This is possible related to the fact that the HM approach involves three classification stages and, if there is a misclassified element at any point, the predicted label is misclassified. In order to overcome this problem we have developed the OsHM approach; which, as mentioned, consists of a combination of the OsM and HM methods that allows generating discriminative features while favoring a straightforward classification. In fact, the OsHM approach yielded a better performance than the HM method for almost all subjects, however, it does not outperform the OsM either. In this sense, even when the classes formulated by the four instances involving the use/not-use of the feet according to the last stage of the HM approach are well separated (see Figure 15), data whose classes were not considered during the models training cause some ambiguities. For instance, in the case of a left hand motor imagery when applying the model both hands-versus-both-hands & feet, decision is not easy since the
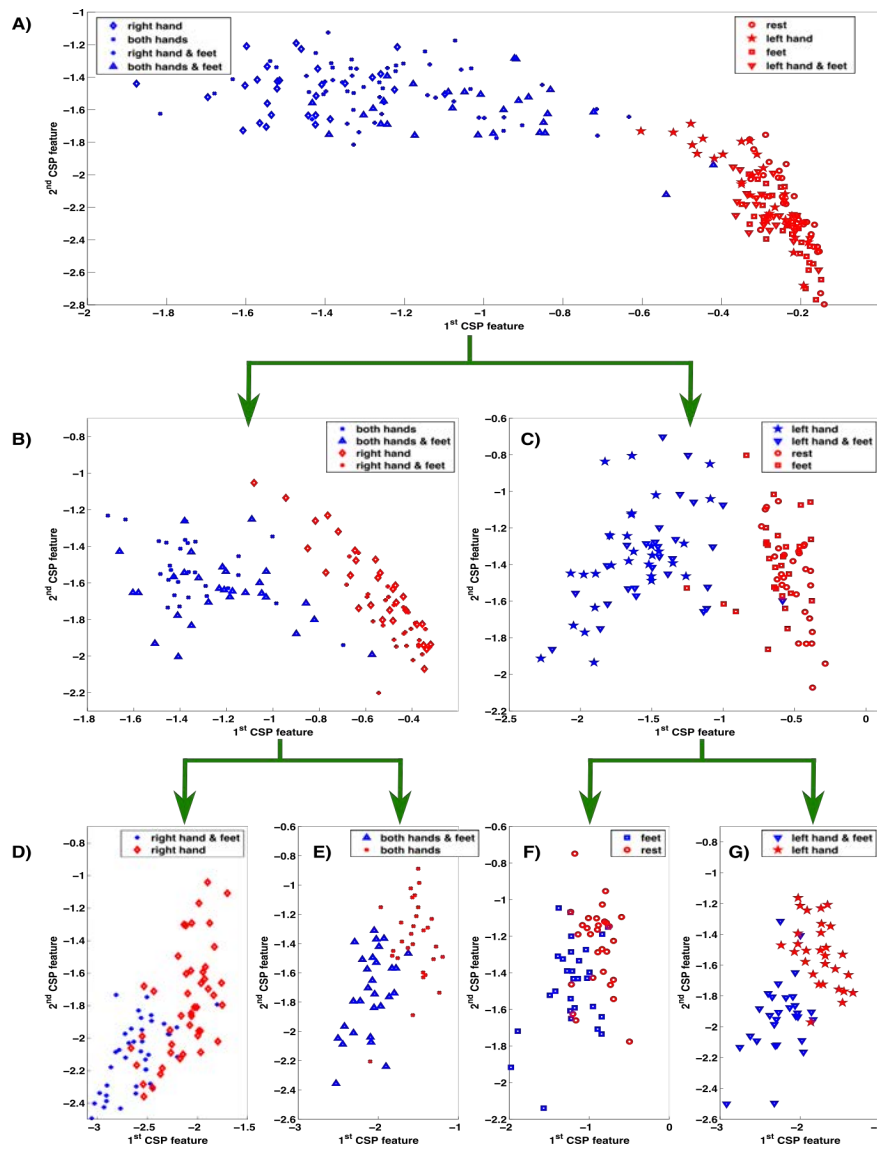
Fig. 14: Distribution of the 2 first CSP features generated by the MH approach on data from subject 3. From top to bottom, the three stages of the hierarchical classification process are shown by starting the process considering the presence/absence of ERD activity related to the use of the right hand. (A) The eight motor tasks of the problem are considered and gathered into two groups: motor tasks that involve the use of the right hand (blue), and motor tasks that do not involve it (red). For the second stage, two models are trained to determine the presence/absence of ERD activity related to the use of the left hand, given the prediction from the previous stage: (B) Motor tasks involving the right hand that also involve the use of the left hand (blue), and motor tasks involving the use of the right hand that do not involve the left hand (red). (C) Motor tasks that do not involve the use of the right hand but that involve the use of the left hand (blue), and motor tasks that do not involve the use of any hand (red). For the third stage there are only two classes remaining and four possibilities given the previous decisions: (D) Rest (red) versus feet (blue); (E) left hand (red) versus left hand & feet (blue); (F) right hand (red) versus right hand & feet, and (G) both hands (blue) versus both hands & feet (red).

left hand is engaged in both tasks. These ambiguities occur within each one of the four models, which presumably hinders an effective classification.

The distribution of the first 2 CSP features generated by the models defined by OsHM approach for evaluating the activity associated to the use of the feet is shown in Figure 15. Note that only the features associated to the feet are presented, since the ones for the left and right hands correspond with the same ones that were already mentioned for the OsM method (see Figure 11). In the same way, the corresponding CSP patterns are, in the case of the hands, those that were presented for the OsM method (see Figure 10-A-C) and, in the case of the feet, the four instances that were described for the HM approach (see Figure 13-D-E-F-G or Figure 12-D-E-F-G, instinctively).

Additionally, in order to compare the specificity of the one-versus-one and the OsM methods in combination with the ACSP algorithm and LDA classification (which had led to the best performance), a graphical representation of the confusion matrices generated by both approaches is presented for all subjects (see Figures 16-26). Note that the surface of each element represents the sum of predictions generating the same label for the trials of each class. The distribution of ambiguities is similar along both methods, however, it is possible to appreciate the slightly superiority of the OsM approach for almost all subjects. Also note that labels are assigned according to the nomenclature discussed in section 5.2 (see Figure 5.4 in page 85).
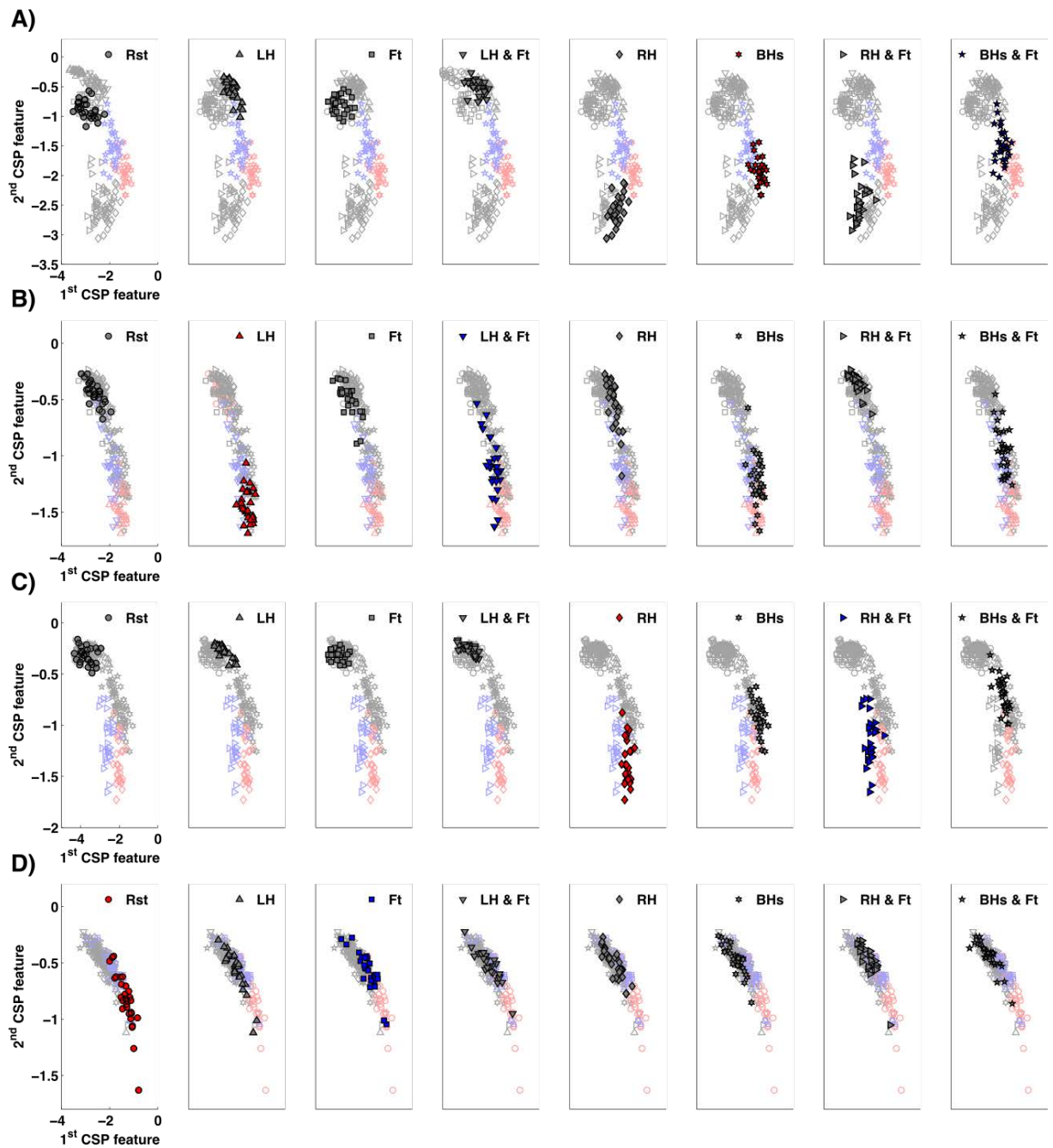
Fig. 15: Distribution of the two first CSP features generated by the OsHM approach for the four instances involving the feet motor task applied on data from subject 3. From top to bottom, the binary problems: both hands-versus-both-hands & feet (A); left hand-versus-left hand & feet (B); right hand-versus-right hand & feet (C), and rest-versus-feet (D). The classes that were used to train the corresponding model are highlighted in blue (for MIs involving the source associated with the feet) and red (for MIs that do not involve the use of the feet). Data that are evaluated by a model that did not consider the corresponding class during the training phase are coloured in gray.

Subject 1

One-versus-One   **One-step Multilabel**
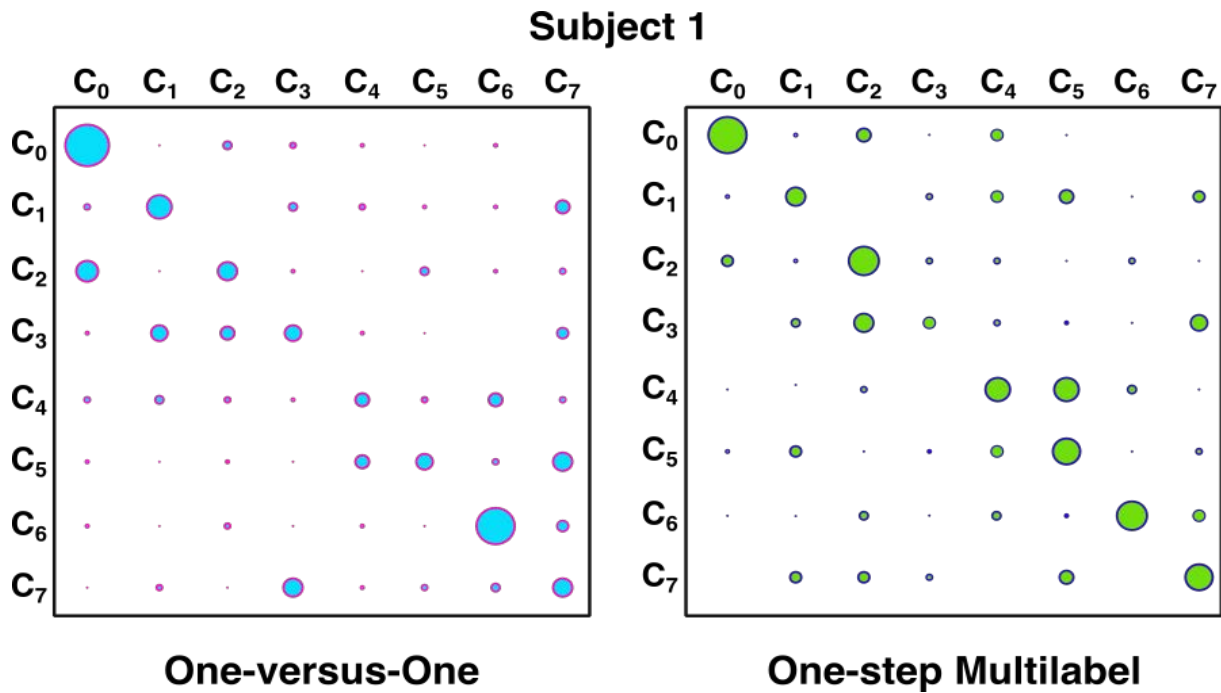
Fig. 16: Confusion matrices obtained by the one-versus-one and One-step Multilabel approaches for subject 1. Note that the surface of each element represents the sum of predictions generating the same label for the trials of each class.

Subject 2

One-versus-One   **One-step Multilabel**
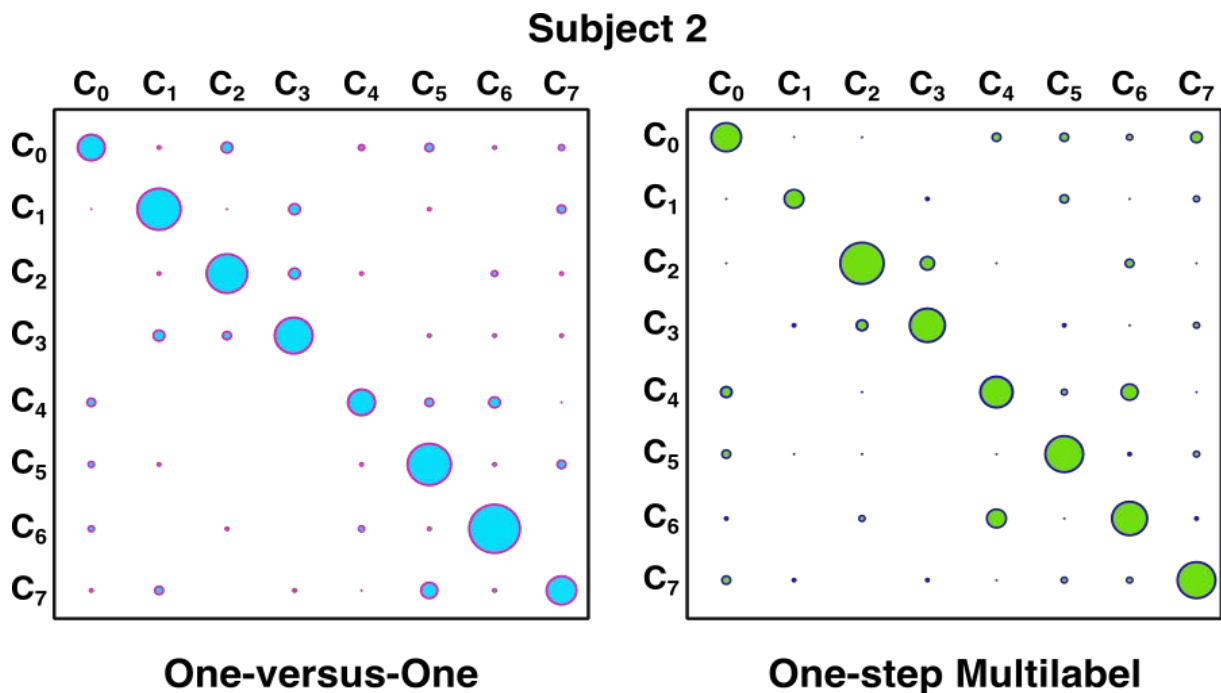
Fig. 17: Confusion matrices obtained by the one-versus-one and One-step Multilabel approaches for subject 2. Note that the surface of each element represents the sum of predictions generating the same label for the trials of each class.

Subject 3

One-versus-One   **One-step Multilabel**
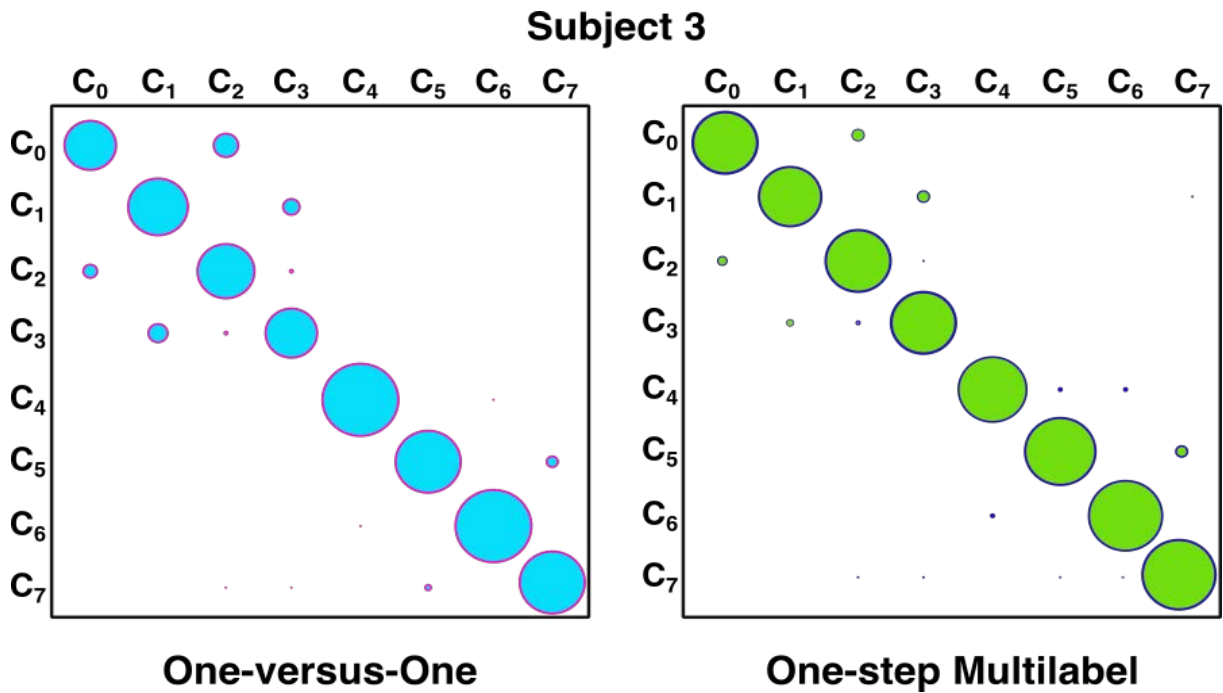
Fig. 18: Confusion matrices obtained by the one-versus-one and One-step Multilabel approaches for subject 3. Note that the surface of each element represents the sum of predictions generating the same label for the trials of each class.

Subject 4

One-versus-One   **One-step Multilabel**
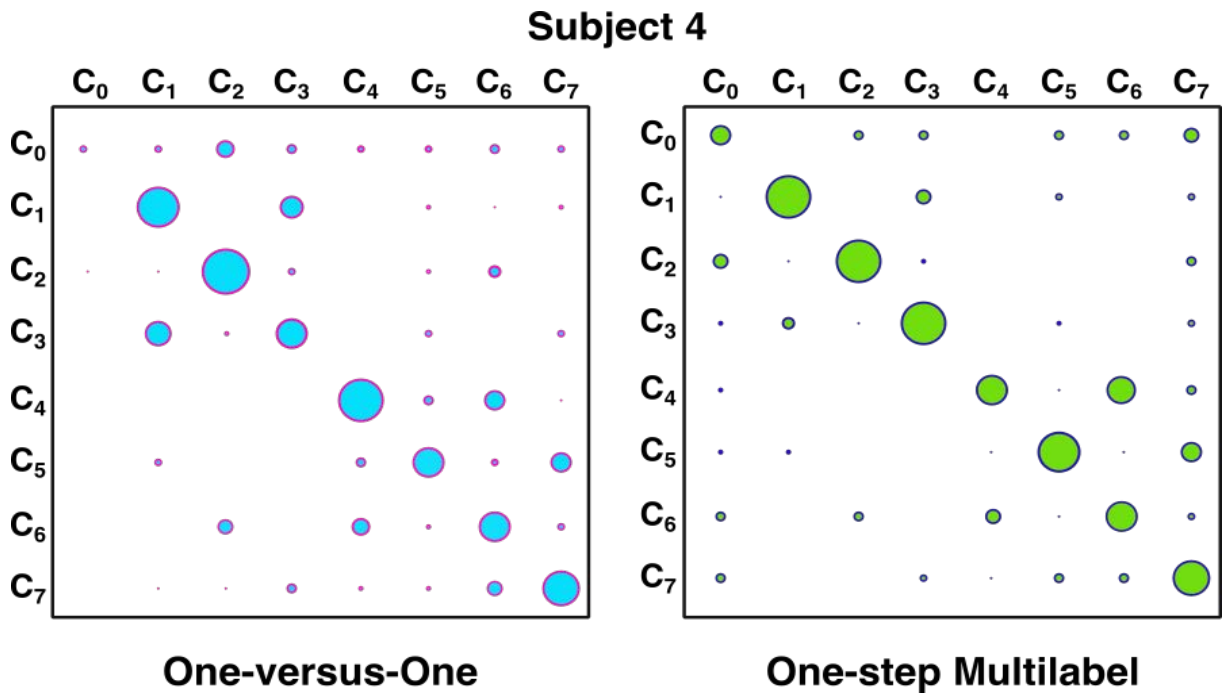
Fig. 19: Confusion matrices obtained by the one-versus-one and One-step Multilabel approaches for subject 4. Note that the surface of each element represents the sum of predictions generating the same label for the trials of each class.

113
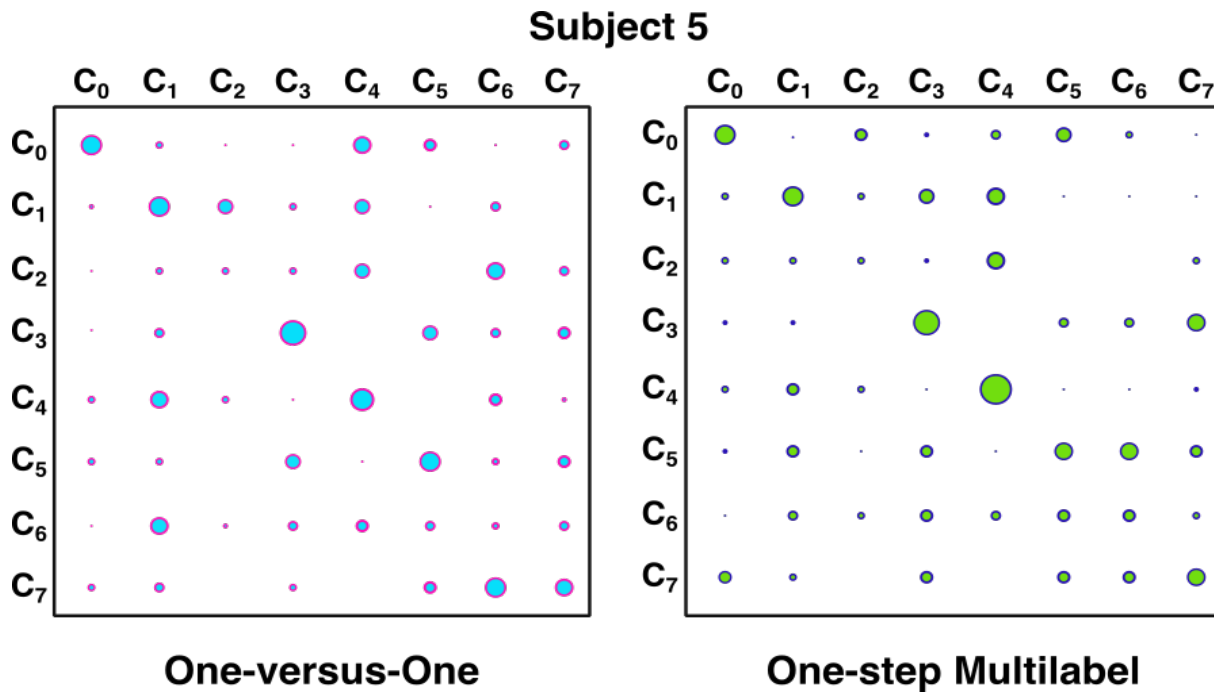
Fig. 20: Confusion matrices obtained by the one-versus-one and One-step Multilabel approaches for subject 5. Note that the surface of each element represents the sum of predictions generating the same label for the trials of each class.
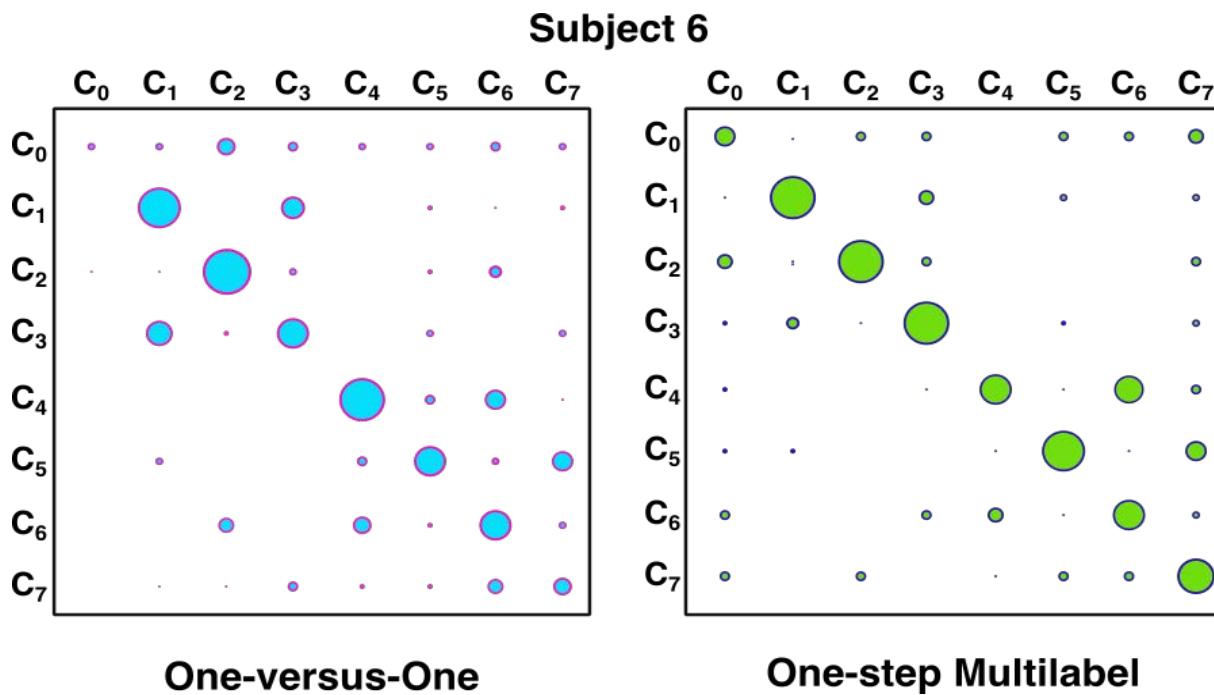
Fig. 21: Confusion matrices obtained by the one-versus-one and One-step Multilabel approaches for subject 6. Note that the surface of each element represents the sum of predictions generating the same label for the trials of each class.
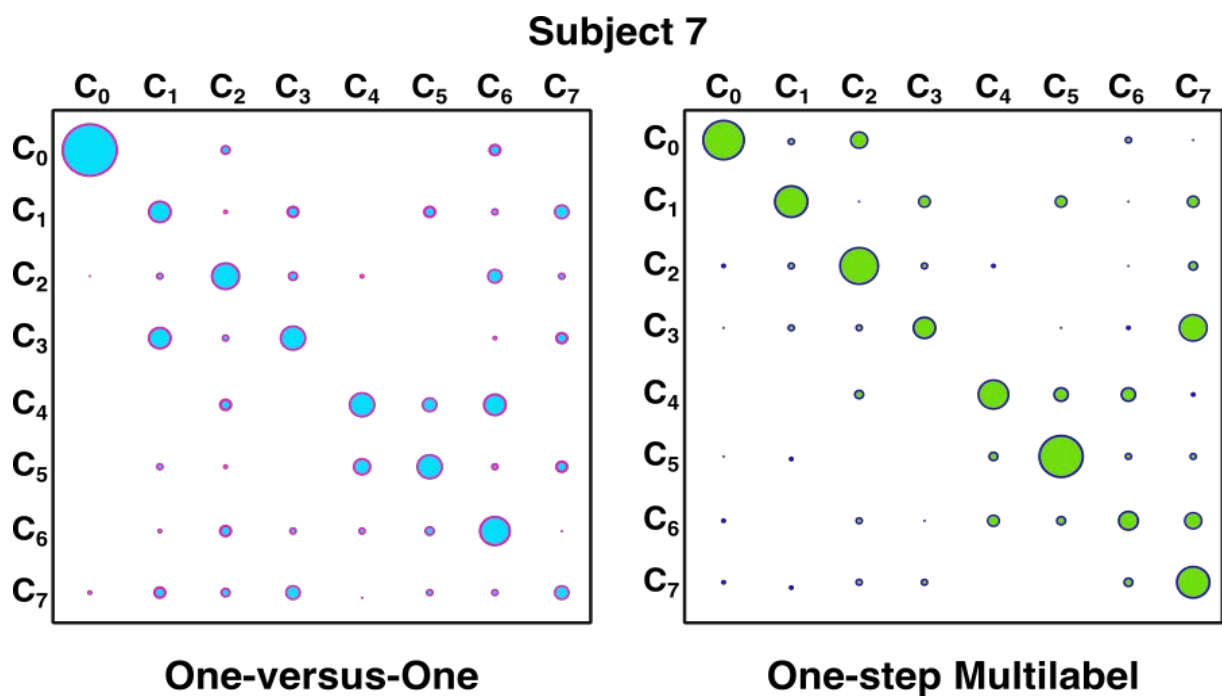
Fig. 22: Confusion matrices obtained by the one-versus-one and One-step Multilabel approaches for subject 7. Note that the surface of each element represents the sum of predictions generating the same label for the trials of each class.
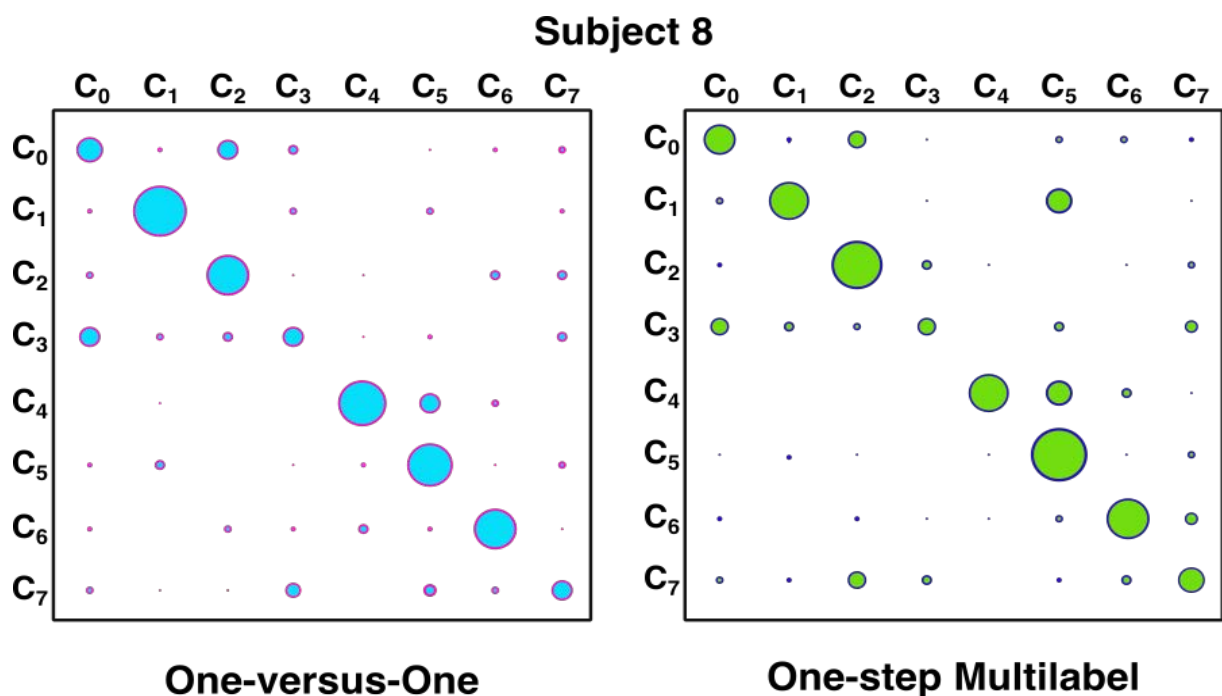
Fig. 23: Confusion matrices obtained by the one-versus-one and One-step Multilabel approaches for subject 8. Note that the surface of each element represents the sum of predictions generating the same label for the trials of each class.
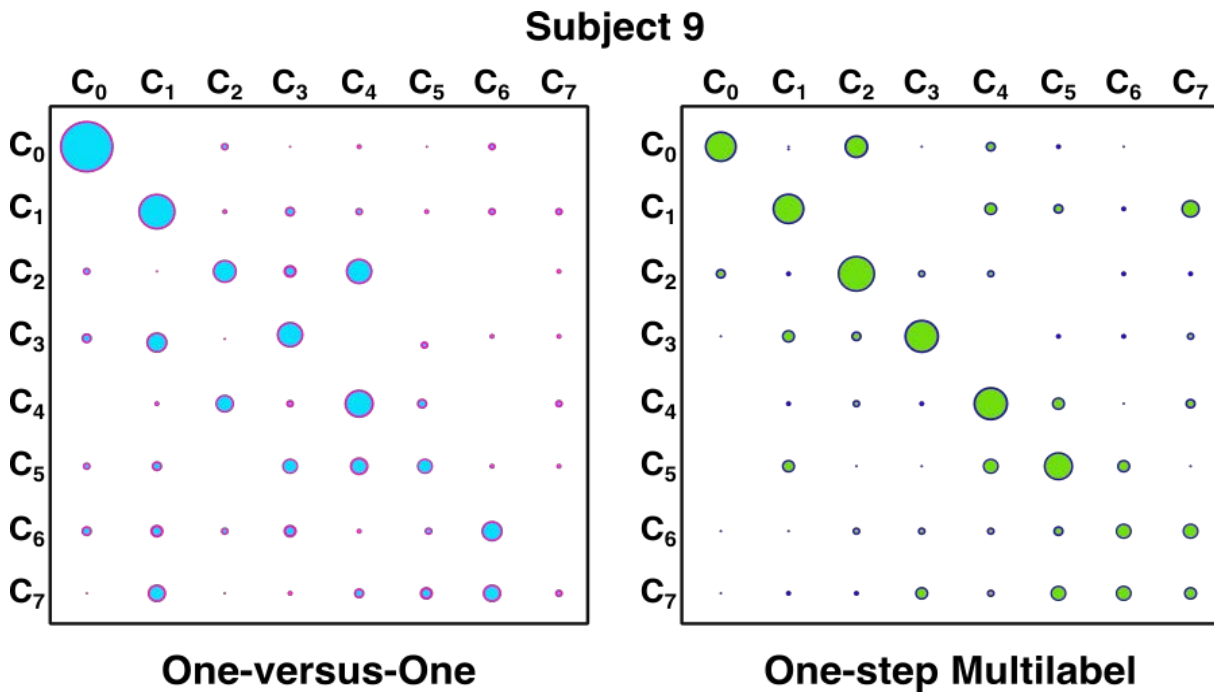
Fig. 24: Confusion matrices obtained by the one-versus-one and One-step Multilabel approaches for subject 9. Note that the surface of each element represents the sum of predictions generating the same label for the trials of each class.
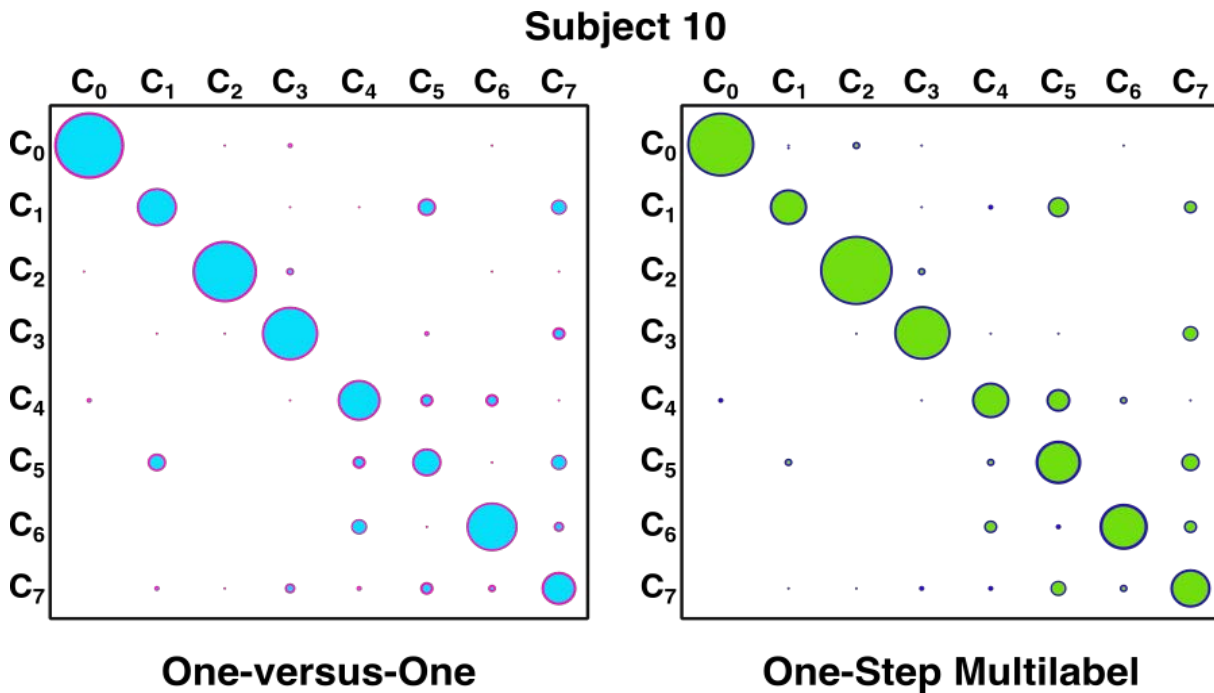
Fig. 25: Confusion matrices obtained by the one-versus-one and One-step Multilabel approaches for subject 10. Note that the surface of each element represents the sum of predictions generating the same label for the trials of each class.
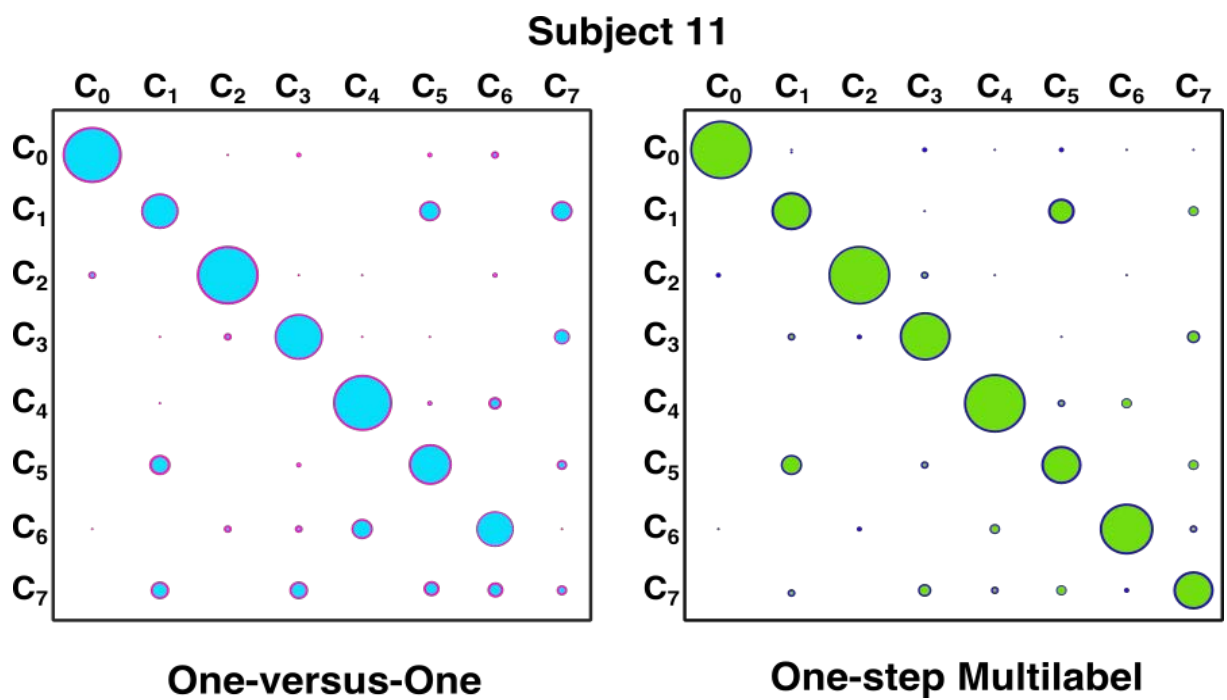
Fig. 26: Confusion matrices obtained by the one-versus-one and One-step Multilabel approaches for subject 11. Note that the surface of each element represents the sum of predictions generating the same label for the trials of each class.

Conclusions

Contents

The present work introduces a new framework for EEG motor imagery-based BCI systems that affords users with an unusual high number of commands, which facilitates a full 3D interaction that can be used in different applications. In this regard, we have implemented a system for controlling a robotic arm that provides fourteen predefined motions.

Such a framework considers the use of combined motor tasks; which contrarily to standard single-label paradigms has not been extensively studied, and contributes with three different methods to solve the associated multiclass problem, namely, One-step Multilabel (OsM) approach, Hierarchical Multilabel (HM) approach, and One-step Hierarchical Multilabel (OsHM) approach, which to our best knowledge have not been considered before. These contributions are based on a particular way of grouping classes that has been inspired by some physiological considerations, such as the sensorimotor rhythms lateralization and the specific location of the activity sources along the sensorimotor cortex. This new scheme allows to transform a multilabel classification problem into a series of binary tasks, which optimizes the use of the Common Spatial Pattern (CSP) algorithm, given that it has been shown to be powerful at discriminating sensorimotor rhythms but has the inconvenience of being suitable for solving only 2-class problems.

We have used two databases in order to analyze the patterns of combined motor tasks, and to implement and validate the proposed methods, namely, the 4-*class* and the 8-*class* databases, which have been specifically recorded for this study. The former database involves only the use of the left and right hands; which together with the rest condition, enable four different brain states when both the single tasks and the combined one are considered (i.e., rest, left hand, right hand and both hands together). In the case of the 8-*class* database, we have also incorporated the use

of the feet, from which eight different brain states are obtained when all possible combinations, including the rest condition, are considered (i.e., rest, left hand, feet, left hand in combination with feet, right hand, both hands together, right hand in combination with feet, and both hands in combination with feet).

The main hypothesis upon which the proposed grouping of classes has been established states that, EEG data recorded during a combined motor imagery, can be characterized as the superposition of the activity that is independently generated by the sources associated to the body parts engaged in the task. In this way it is possible to generate a series of binary classification problems in which the activity induced by the main sources is considered separately, so that one class is conformed by all mental states inducing ERD activity within the corresponding source (i.e., all motor imageries involving the use of the associated limb), and the other class by the mental states that do not induce ERD activity over the associated source (i.e., all motor imageries that do not involve the use of the associated limb). This hypothesis has been motivated by the distribution of the relative oscillatory power that is observed over the primary motor cortex during the different motor imageries; which, as can be seen from figures 2.9 in page 29 and 5.4 in page 85, appears in the contralateral hemisphere of the body parts that are involved during the different motor tasks. Moreover, we have presented a series of statical analysis to confirm the plausibility of such assumption (see figure 2.15 in page 35).

By using this scheme, we have applied the CSP algorithm to solve the binary problems associated to each activity source. In this way, it is possible to directly apply the CSP algorithm without having to resort to any multiclass extension, such as the one-versus-one or one-versus-all approaches, and hence to optimize its performance. The first method that we have presented, namely, One-step Multilabel (OsM) approach, directly applies this solution, so that P binary problems, where P corresponds to the number of body parts considered by the paradigm, are formulated (2 binary problems for the 4-*class* database and 3 binary problems for the 8-*class* database). Subsequently, the extracted features are concatenated and used to train/validate a classification model.

The two other contributed algorithms, namely, Hierarchical Multilabel (HM) and One-step Hierarchical Multilabel (OsHM) methods, consist of two variants of the OsM approach. These two methods were motivated by the poor discrimination that was observed over the features generated by the binary problem associated to the central activity source (i.e., related to the feet) after applying the first approach on the 8-*class* database. As mentioned, this deficiency to well separate the corresponding elements is presumably caused by the fact that the central activity source is located too close from the sources associated with the hands, which influences its behavior. By considering this aspect, the two remaining methods have been conceived in such a way that the main decisions are based on the activity elicited by the sources related to the hands and, on a complementary level, by the activity that is linked to the feet. Thus, the HM method consists of a hierarchical approach for which the first stage of the classification involves a binary problem comprising in one class all motor imageries involving the use of one hand and, in

the second class, the motor imageries that do not include it. Once the prediction is made, half of the classes are discarded and the process continues in the second stage where, according to the group that was previously chosen, one of two models is selected. Each one of these models consist again in a binary problem using the same type of partition: from the remaining classes and given that the hand that was already analyzed was found to be active or inactive in terms of motor imagery-related activity, one class is conformed by the motor tasks involving the use of the second hand and, the other class, by the motor tasks that do not include it. Finally, once the prediction for the second stage is made, there are only 2 classes left; which present the same activity associated with the left and right hands but differ in the activity that is linked to the feet. In this way, by considering only 2 classes, the task associated with the feet is simplified and addressed by a third and last stage. In this regard, it must be noticed that the HM method stops in the second stage when it is applied on the 4-*class* database since the feet were not included in the paradigm. It should be also noticed that the HM method can be applied by considering in the first stage either the left hand or the right hand. In this sense, there was a slightly improvement when starting with the right side, though it is not really significant. Even when separation was better in the case of the 8-*class* database for the binary problems involving the feet, results remain below that the ones achieved by the OsM and OsHM methods. This might be presumably due to the fact that, if there is a misclassified element at any stage, the predicted label will be wrong, whereas for the other two methods, classification is carried out by a single model over the concatenation of features.

Finally, the OsHM algorithm results from the combination of the two previous methods; which consists in taking the 2 binary problems that were introduced for each hand by the OsM method, and incorporates the 4 models that were formulated by the HM approach for the analysis of the activity generated by the feet. All features are computed independently and simultaneously, and the resulting elements are concatenated in order to train/validate a classification model. The OsHM approach was applied only on the 8-*class* database since in the case of the 4-*class* one it is equivalent to the OsM method.

The best grand average performance across subjects for the 8-*class* database was obtained by the OsM method in combination with the ACSP algorithm, which is followed be the OsHM in combination with the CSP algorithm. As mentioned in the discussion, even though the OsHM approach generates more discriminative features than the OsM method by formulating the four instances involving the use/not-use of the feet in relation to the engagement of the hands, some ambiguities are caused by data whose classes were not considered during the models training, which hinders an effective classification. However, both approaches are slightly better than those generated by the classical methods, which is of particular interest if we consider that, in order to solve the same problem, the one-versus-one and one-versus-all approaches require respectively, 28 and 8 feature extraction models; which is by far more than the 3, and 6 that are needed, respectively, by the OsM and OsHM approaches. In the case of the 4-*class* database the best results were obtained by the HM approach in combination with FBCSP algorithm and linear

discriminant analysis for classification. In this sense, since the problem involves only two classification stages, the risk of failure is not as high as in the case of the 8-*class* database.

Results and the corresponding discussion were presented, respectively, in chapters 6 and 7; where, aiming at comparing the performance yielded by each one of the proposed feature extraction methods, we have also included the accuracy generated by the classical one-versus-one and one-versus-all approaches, as well as the one obtained by the CSP by Joint Diagonalization (CSP by JAD) algorithm. In order to explore different solutions for all cases involving the use of the CSP algorithm, we have presented the results generated by two of its variants, namely, ACSP and FBCSP. Moreover, we have also reported the performance of the Minimum Distance to Riemannian Mean (MDRM) algorithm, which has been drawing the attention of the BCI community due to its capability for discriminating sensorimotor rhythms. In this regard, we have contributed with a new approach that uses the CSP method in order to increase the contrast between two conditions before mapping their covariance matrices into the Riemannian space, which even when it does not outperform the CSP approach, yields better results than the MDRM applied independently, showing that it is advantageous to use the CSP algorithm as a preprocessing method.

Interestingly, for classification results were always better by using linear discriminant analysis than support vector machines; presumably because the latter approach is more likely to get overtrained.

To conclude, we have proposed three new classification methods that are capable of distinguishing an unusually high number of motor imageries to enable the full 3D control of a robotic arm by using EEG. Under the proposed control scheme, the system manages to map 14 different movements by using only three body parts. Results show a varying accuracy depending on the subject, but it must be considered that the amount of data was limited and that the grand average performance of two of the proposed methods exceeded the classical one-versus-one and one-versus-all approaches.

8.1   Contributions

The main contribution of the present work is the evidence that EEG data recorded during a combined motor imagery can be modelled as the superposition of the activity that is independently generated by the sources associated to the body parts engaged in the task. This assumption can be exploited in different ways, as we have done by proposing an innovative way of grouping classes to address a multilabel paradigm. From this scheme, we have presented three new methods that not only optimize the feature extraction method by allowing to directly use the CSP algorithm on different binary problems, but they also reduce the complexity of the classification task, since contrarily to the classical feature extraction approaches based on the one-versus-one and one-versus-all algorithms, for which different features are required to train each one of the involved classifiers, the proposed methods use the same over all models, which has a positive impact on

the computation time.

Besides the three feature extraction methods that have been proposed, we have also explored and contributed with different solutions to the standard CSP algorithm. As in the case of the new approach presented in section 4.6.2.2, which uses the spatial filters generated by the CSP algorithm as a preprocessing step before mapping the covariance matrices of the CSP projected trials into the Riemannian manifold, from where the class labels of unseen elements can be assigned by using the Minimum Distance to Riemannian Mean (MDRM) algorithm. Even though the performance of such approach was not better than the one achieved by the standard CSP algorithm, it outperforms in all cases the one that is obtained by directly applying the MDRM method, showing that it is advantageous to merge both approaches. The justification of this behaviour is presumably given by the fact that, the spatial filters generated with the CSP algorithm extract information that enhances the difference between two conditions; which is preserved by the covariance matrices of the projected trials. Since such covariance matrices serve as descriptors, from where it is possible to extract the spatial information, such enhancement is reflected in the method performance.

Furthermore, we have also implemented some variants of well-known methods that, to our best knowledge, have not been reported before; such as the Filter Bank Analytical Common Spatial Pattern (FBACSP) algorithm, and a hierarchical approach based on the Minimum Distance to Riemannian Mean (MDRM) method [109].

Finally, another important aspect to consider from the implementation design, is the switching-mode control, which allows employing the same commands to carry out different actions. Under this scheme it is possible to manage an even higher number of commands, from which it is possible to afford users which a full 3D control.

8.2   Limitations

The principal limitation of the proposed methods lies in their major attribute; which consists of the way how classes are grouped to generate the binary problems associated to each activity source. While this may optimize performance and computation time for users whose EEG patterns present a strong lateralization, it may not be effective for those subjects whose EEG activity lacks this differentiation. However, most of the feature extraction methods for sensorimotor rhythms discrimination are based on such a contrast, so this issue does not really represent an inherent drawback of the proposed approach, but a reality that has to be faced by BCI researchers.

Another limitation related to the operation scheme is the switching-mode function; which has the benefit of drastically increasing the number of available commands, but presents the drawback of slowing down the control of the device. In this regard, it is reasonable to consider imposing a particular assignment of classes that associates the most discriminative ones to the most frequent commands. In this way, misclassification rate of common functions, such as the switching component, would be reduced and thus efficiency enhanced.

In general, the proposed implementations share the same complications and problems as any motor imagery based EEG BCI system. For instance, one of the biggest challenges is to understand and solve the problem of BCI-Illiteracy, which implies that a substantial portion of users (estimated within 15% and 30%) do not manage to achieve control on this kind of systems [115]. This condition appears in different levels and, for some users, it is possible to generate a classification model during the calibration phase, although results are unsatisfactory when it is applied to provide online feedback. Whereas, in other cases, it is not even possible to generate a classification model during the training stage. Besides the illiteracy problem, motor imagery based BCIs are susceptible to intra-session variability, which provokes inconstancy of data recorded from the same user throughout different dates as a result of possible changes in the electrodes location, and/or variations of the mental state of the subject. This variability can cause that the initial calibration is not longer accurate for subsequent settings, which encumbers the daily use of the system. Also in this regard, it should be noticed that performance in out-of-the-lab environments gets affected by additional noise sources, as well as by different types of distractions that may cause non-stationary features [116].

Finally, the most important challenge for adopting this kind of applications may lie in the establishment of effective training methods, since the existing protocols essentially provide theoretical considerations that do not stimulate the acquirement of appropriate skills to effectively modulate EEG patterns, which might be responsible for inadequate user performance and BCI illiteracy [117].

8.3   Future work

The main interest is to implement the proposed solution to enable an online control, which will provide users with a self-paced operation mode affording them with the autonomy to carry out actions when desired (i.e., an asynchronous protocol). To this end there are two aspects to be considered: the online detection of the different commands, and the duration of each motion. For the first consideration, either the OsM approach or the OsHM method can be applied by using a sliding window, and thus to detect at each step the motor task that the user is performing. For the second aspect, we propose to implement the post-movement rebound detection as we have presented in [118]. In this way, users will have more control to direct the device according to their needs; contrary to a predefined motion duration, where the users goal could lie in-between the range of movement. Moreover, regardless the application that is attempted to be controlled by using the proposed implementation, there are some considerations that can help to facilitate the classification task. For instance, in the case of the robotic arm, as it was discussed in section D.1, there is a limited range of movements once the device has reached an extreme position, so that even if the system detects a command demanding a motion that is beyond the possibilities of movement, the corresponding action will not be executed. In this sense, an effective approach will be to reduce the classification task in such a way that only the mental states related to a suitable

command are considered, so that less classes are evaluated and thus the classification task is facilitated, which would result in an improvement of the system performance. Another aspect that may improve the system performance consists in using users reaction upon a predicted action as feedback to continuously calibrate the system. It is well-known that a subject's recognition of having committed a response error is accompanied by specific variations that can easily be observed in averaged event-related potentials (ERP) [119], which can be used in many ways either for avoiding false positive actions, or for generating adaptive classification models. As part of the feature work, it remains to explore this solution in order to provide an adjustable system.

As mentioned, a very important aspect that has to be considered in order to develop effective applications consists in conceiving adequate training techniques. While important progress has been made on this field [117], the use of combined movements demands an even higher competence than the required over simple motor imageries; which has to be address if satisfactory performance is to be achieved. In this sense, it is interesting to note that the proposed methods favor a sensorimotor rhythms discrimination that is based on their lateralization, which represents one of the main physiological characteristics of such mental task. In this regard, such scheme for feature extraction might be more suitable than the one-versus-one approach for providing feedback during users' training, given than the resulting improvement would be based on the achievement of this specific differentiation, which is directly related to the metal process.

Finally, as mentioned in section 6.1, there are some subject-specific experimental parameters that can be tuned to increase the system performance, namely, the frequency for band-pass filtering of the EEG signals, the time window of the EEG trials taken relative to the stimuli, and the subset of CSP filters to be used [94]. In this sense, it remains to integrate optimization algorithms to find the best possible adjustment of such parameters working in combination with the proposed methods, and thus to boost the general performance of the system. In the same way, regularization of the classification algorithms would be also pertinent.

Appendices

A

Résumé Étendu

Les interfaces cerveau-ordinateur (BCI) substituent aux sorties naturelles du système nerveux des sorties artificielles qui n'exigent pas l'utilisation des nerfs périphériques. Ceci permet à des patients atteints de troubles moteurs sévères d'interagir, en utilisant seulement leur activité cérébrale, avec différents types d'applications: correcteur orthographique, neuroprothèse, fauteuil roulant, ou dispositifs robotiques [1], [2]. L'utilisation des interfaces cerveau ordinateur n'est pas restreinte, et elles ont également suscité de l'intérêt pour le développement de programmes de réadaptation destinés à des patients atteints d'AVC [3], [4], ainsi que des interfaces pour les personnes en bonne santé (par exemple, dans les applications de jeu).

Plusieurs types de signaux cérébraux peuvent être utilisés pour le contrôle BCI. L'un d'entre eux, particulièrement intéressant puisqu'il a l'avantage de favoriser les paradigmes auto-stimulés (c'est-à-dire un mode de fonctionnement asynchrone), est connu sous le nom d'imagerie motrice (IM); ceci peut apporter un contrôle intuitif et naturel pour permettre à des personnes à mobilité réduite une interaction riche avec leur environnement. L'imagerie motrice implique la pratique mentale d'une tâche motrice sans effectuer de mouvement apparent [5]. L'imagination d'un mouvement implique les mêmes régions du cerveau, et des fonctions similaires à celles engagées dans la programmation et l'exécution d'un mouvement réel. La principale différence est que dans le premier cas, les mouvements sont bloqués au niveau cortico-spinal [6].

L'activité EEG obtenue lors de l'imagerie motrice consiste en des changements de rythmes sensori-moteurs qui présentent un comportement oscillatoire particulièrement intéressant dans deux gammes de fréquences spécifiques: les bandes alpha (8-13 Hz) et beta (13-25 Hz) [7]–[9], puisqu'il a été montré qu'ils sont modulés pendant et après la préparation et l'exécution des

mouvements imaginés [10]. Cette modulation prend deux formes différentes: une réduction de puissance oscillatoire au cours de la tâche motrice appelée désynchronisation événementielle (ERD), et une augmentation de puissance oscillatoire connue sous le nom de Event-Related Synchronization (ERS) qui apparaît juste après la fin de la tâche motrice, et qui dure quelques centaines de millisecondes [11]. Lors de la préparation d'une tâche motrice, certaines régions corticales et sous-corticales sont activées. A l'inverse, au cours de l'imagerie motrice, la plus grande partie de l'activité a lieu dans le cortex moteur principal, dans l'hémisphère contralatéral correspondant [12]. On sait que la représentation topographique et le changement de puissance de la bande au cours de l'imagerie motrice sont localisés dans des régions spécifiques du cortex moteur primaire (en gros, les activités correspondant au côté gauche sont observées sur l'électrode C4, les activités correspondant au côté droit sur C3, les pieds sur C_z, etc.) [11], [13].

Le nombre de commandes que permet un système BCI à base d'imagerie motrice dépend du nombre d'états cérébraux que le système est capable de détecter, quantité qui est limitée par le nombre de parties du corps avec lesquelles les utilisateurs sont capables de produire des modèles d'activité appropriés. De plus, la difficulté de la tâche augmente avec le nombre total de commandes disponibles. Augmenter le nombre de commandes n'est possible que si l'on peut discriminer de manière efficace un nombre correspondant d'états mentaux [14]. Il existe deux types d'approches pour résoudre des problèmes multiclasses: D'une part, certaines techniques tentent de prédire le label en utilisant une méthode directe, comme les algorithmes basés sur la distance. D'autres part, certaines méthodes utilisent des extensions multiclasses afin de réduire le problème à un ensemble de décisions binaires qui peuvent être accomplies de différentes façons, par exemple, en utilisant les approches classiques one-versus-one ou one-versus-all [15].

Une méthode multiclasse intéressante, qui permet aux utilisateurs de disposer d'un vaste répertoire de commandes différentes, est connue sous le nom d'approche multi-label. Cette approche est basée sur la détection d'imageries motrices combinées (c'est-à-dire impliquant deux ou plusieurs parties du corps à la fois) [16], [17]. En comparaison avec le schéma standard, cette approche présente l'avantage d'augmenter considérablement le nombre d'états différents du cerveau détectables, tout en utilisant le même nombre de parties du corps (de l'ordre de 2^P par rapport à P, où P est le nombre de parties du corps impliquées).

Contrairement aux paradigmes multi-classes standard, auxquels laplupart des recherches ont été consacrées, moins de travail a été rapporté sur les modèles oscillatoires cérébraux induits par les imageries motrices combinées. La technique de décomposition en mode empirique combinée avec l'analyse de synchronisation de phase a été proposée et appliquée pour évaluer la synchronisation des zones cérébrales motrices supplémentaire et primaire lors d'une imagerie motrice complexe combinant l'action du corps et des membres. Les résultats de classification ont suggéré la faisabilité de cette approche pour des applications BCI de récupération motrice des membres inférieurs [18]. En outre, afin de surmonter la quantité limitée de commandes de sortie résultant du nombre limité de classes, et dans le but d'obtenir un contrôle 3D continu, l'imagerie motrice des deux mains a été adoptée afin d'augmenter le jeu d'instructions d'une interface BCI basée sur l'imagerie

motrice d'un seul membre [17]. Plus récemment, on a étudié les caractéristiques des modèles EEG au cours d'imagerie motrice simple et combinée, afin de confirmer la séparation de sept classes différentes pour la mise en œuvre des systèmes BCI à base d'imagerie motrice [16]. Une autre étude a rapporté que le mouvement de plusieurs doigts en même temps n'est pas la simple superposition de mouvements de doigt individuels à l'échelle ECoG, mais plutôt l'activation de réseaux de neurones supplémentaires. Ceci implique que pour mettre en œuvre une interface BCI, la génération de signaux de contrôle codant des mouvements complexes devrait probablement utiliser des composantes de signal d'ordre plus élevé que la superposition des signaux individuels des membres [19]. Cependant, à notre connaissance, cette dernière conclusion reste à confirmer à l'échelle EEG; ceci est la question principale de notre travail.

L'importance de cette question découle de l'intérêt porté à optimiser l'algorithme Common Spatial Pattern (CSP) pour l'appliquer à des tâches de classification multi-label. Cette méthode est un algorithme très pratique et beaucoup utilisé pour extraire des caractéristiques d'imagerie motrice à partir de signaux EEG. Il génère une série de filtres spatiaux qui sont appliqués pour décomposer des données multidimensionnelles en un ensemble de composantes non corrélés [20], qui visent à extraire des éléments qui maximisent simultanément la variance d'une classe tout en minimisant la variance de l'autre. En raison de la manière dont l'algorithme CSP réalise un tel contraste, cette approche présente l'inconvénient de ne convenir que pour discriminer entre deux classes et, pour résoudre des problèmes impliquant de nombreuses conditions, des extensions multi-classes sont nécessaires [2], [15]. Les extensions classiques telles que les approches one-versus-one et one-versus-all présentent plusieurs inconvénients. Peut-être le plus important est la quantité de modèles qui sont nécessaires à mesure que le nombre de classes augmente, ce qui a un impact négatif sur le temps de calcul et sur la fiabilité des résultats. En ce sens, si nous pouvons confirmer que l'imagerie motrice d'une combinaison de deux ou plusieurs parties du corps impliquées en même temps correspond à une simple superposition des membres individuels à l'échelle EEG, alors il sera possible de concevoir des approches plus simples pour adapter non seulement la méthode CSP mais n'importe quelle autre méthode, puisque selon ce schéma, il deviendra possible d'utiliser un ensemble de classificateurs binaires afin de déterminer, pour chaque source d'activité, si la partie corporelle qui lui est associée est ou non engagée dans la tâche et donc inférer la décision finale à partir d'une combinaison de tous les sous-problèmes [21], [22].

Dans la présente étude, nous fournissons quelques éléments de preuve suggérant que cette hypothèse est plausible, sur laquelle nous avons développé trois nouvelles méthodes d'extraction de caractéristiques pour discriminer les imageries motrices combinées, à savoir l'approche multi-label à une étape (OsM), l'approche hiérarchique multi-label (HM) et l'approche hiérarchique multi-calques (OsHM) en une étape, qui à notre connaissance n'ont pas été considérées auparavant. La motivation était de permettre le contrôle 3D complet d'un bras robotique commercial qui offre aux utilisateurs quatorze mouvements prédéfinis. Le paradigme proposé implique l'utilisation de trois parties du corps, (à savoir: main droite, main gauche et pieds), qui en combinaison avec la

condition de repos, constitue un problème à huit classes. Pour le mettre en œuvre, nous avons utilisé une commande qui permet de changer de mode, afin de pouvoir associer différentes actions à la même commande. Six sujets en bonne santé ont participé à une séance d'enregistrement comprenant trois séries de dix essais par classe, soit 30 essais par classe. Les résultats sont légèrement supérieurs à ceux obtenus par les approches classiques one-versus-one et one-versus-all, ce qui est particulièrement intéressant si l'on considère que, pour résoudre le problème de classe 8 susmentionné, les deux méthodes exigent respectivement 28 et 8 modèles d'extraction de caractéristiques, ce qui est beaucoup plus que les 3, 7 et 6 modèles requis par les approches OsM, HM et OsHM respectivement.

Cette thèse est structurée de la manière suivante: Le premier chapitre est consacré à fournir une base générale des principes et des pratiques des systèmes d'interface cerveau-ordinateur (BCI). Ces aspects seront approfondis le long du deuxième chapitre dans le contexte de l'imagerie motrice. Au cours de ces chapitres, nous présenterons la base de données à 4 classes, qui comprend une série d'enregistrements EEG de cinq sujets en bonne santé qui ont participé à une expérience d'imagerie motrice impliquant l'utilisation de la main droite, de la main gauche et des deux mains ensemble; ce qui, en plus de la condition de repos, conduit à un paradigme de quatre classes différentes. La base de données à 4 classes a été créée pour réaliser une étude préliminaire afin d'évaluer les similitudes entre l'activité obtenue sur les électrodes associées à l'utilisation de la main gauche et de la main droite lors d'imagerie motrice simple et combinée. De cette façon, en proposant un petit nombre de classes, il est possible d'enregistrer une quantité considérable d'essais sans provoquer la fatigue des utilisateurs. Sur ces données nous présenterons une série de résultats statistiques basés sur l'analyse du spectre de puissance observé sur chaque électrode au cours des quatre tâches motrices différentes, ce qui étaye l'hypothèse que les données EEG enregistrées au cours d'une imagerie motrice combinée peuvent être caractérisées comme la superposition de l'activité générée indépendamment par chacune des sources associées aux parties du corps engagées dans la tâche. Par la suite, dans le troisième chapitre, nous allons décrire les caractéristiques du bras robotique qui a motivé les solutions pratiques développées tout au long de ce projet. Dans la même ligne, nous allons décrire le schéma d'opération que nous avons conçu, qui consiste en une implémentation intelligente basée sur un contrôle à deux modes, qui permet de gérer différentes actions en utilisant la même commande. Comme mentionné, ce dispositif offre aux utilisateurs quatorze mouvements prédéfinis, ce qui exige beaucoup plus de commandes que celles obtenues avec la base de données à 4 classes. Dans cette optique, nous avons généré une base de données à 8 classes, qui comprend des signaux EEG de 6 sujets sains produisant les huit états mentaux qu'il est possible de générer par les tâches motrices impliquant l'utilisation de la main droite, de la main gauche, des deux pieds ensemble, et le repos. De la même manière que pour la base de données à 4 classes, nous présenterons une analyse qualitative des modèles ERD qui se trouvent dans les signaux EEG au cours de huit tâches différentes, ainsi qu'une comparaison statistique du spectre de puissance sur chaque électrode pour chaque couple de conditions. Ces résultats constituent la base sur laquelle nous justifierons le regroupement

des classes qui a été suggéré pour les méthodes d'extraction de caractéristiques apportées par le présent travail. Dans le quatrième chapitre, le concept d'extraction de caractéristiques est introduit et élaboré dans le contexte des signaux EEG au cours de l'imagerie motrice. A travers ces pages, l'algorithme CSP sera élucidé avec deux de ses variantes, à savoir les modèles spatiaux communs analytiques (ACSP) et les modèles spatiaux communs par Diagonisation Approximative Conjointe (CSP by JAD). De plus, nous présenterons une stratégie appelée Filter Bank Common Spatial Pattern (FBCSP) qui optimise la performance de l'algorithme en décomposant les signaux EEG en une série de bandes de fréquences, ce qui favorise la sélection d'une gamme de fréquences discriminative spécifique au sujet. Par la suite, nous présenterons la notion de classification, pour laquelle deux algorithmes couramment utilisés seront décrits, à savoir l'Analyse discriminante linéaire (LDA) et les Machines à Vecteur de Support (SVM). De plus, afin d'explorer différentes solutions, le présent travail considère aussi l'utilisation de la géométrie riemannienne, avec laquelle la classification peut être obtenue directement dans l'ensemble des matrices symétriques définies positives qui est donné par les matrices de covariance des signaux EEG. Ensuite, dans la dernière section de ce chapitre, nous introduirons une nouvelle méthode de classification des systèmes BCI basée sur l'imagerie motrice, qui utilise la combinaison de l'algorithme CSP et de la géométrie riemannienne, ce qui représente l'une des contributions du présent travail. Dans le cinquième chapitre, le concept d'approche multiclasse sera introduit en décrivant trois méthodes classiques, à savoir les algorithmes un-contre-un, un-contre-tout et hiérarchique. À ce stade, nous décrirons les trois nouvelles méthodes multilabel que nous avons mises en œuvre, à savoir l'approche multi-label à une seule étape (OsM), l'approche hiérarchique multi-label (HM) et l'approche hiérarchique monostable (OsHM) à une étape. Chacune de ces nouvelles méthodes est fondée sur les résultats qui ont été présentés dans les chapitres 2 et 3, à partir desquels nous avons proposé une méthode pratique pour effectuer la partition des données, de sorte que tout le problème est réduit à une série de tâches de classification binaires. Enfin, au long du sixième chapitre, nous présenterons les résultats, en termes de précision, obtenus par chacune des méthodes décrites sur les deux bases de données. La discussion correspondante est présentée dans le chapitre 7 et la conclusion générale au chapitre 8.

B

Time series of ERD/ERS% patterns

Along this section a series of complementary images are presented in order to show the ERD/ERS% patterns that are obtained independently for each subject. Images were obtained by applying the same method described in sections 2.5 and 3.1.2.

For each subject such procedure was applied over the band-passed filter trials over the alpha [7-14 Hz] and alpha+beta [8-30] bands, within which time series were computed along the time period from [0.5-4.5 s] see Figures B.1-B.22)
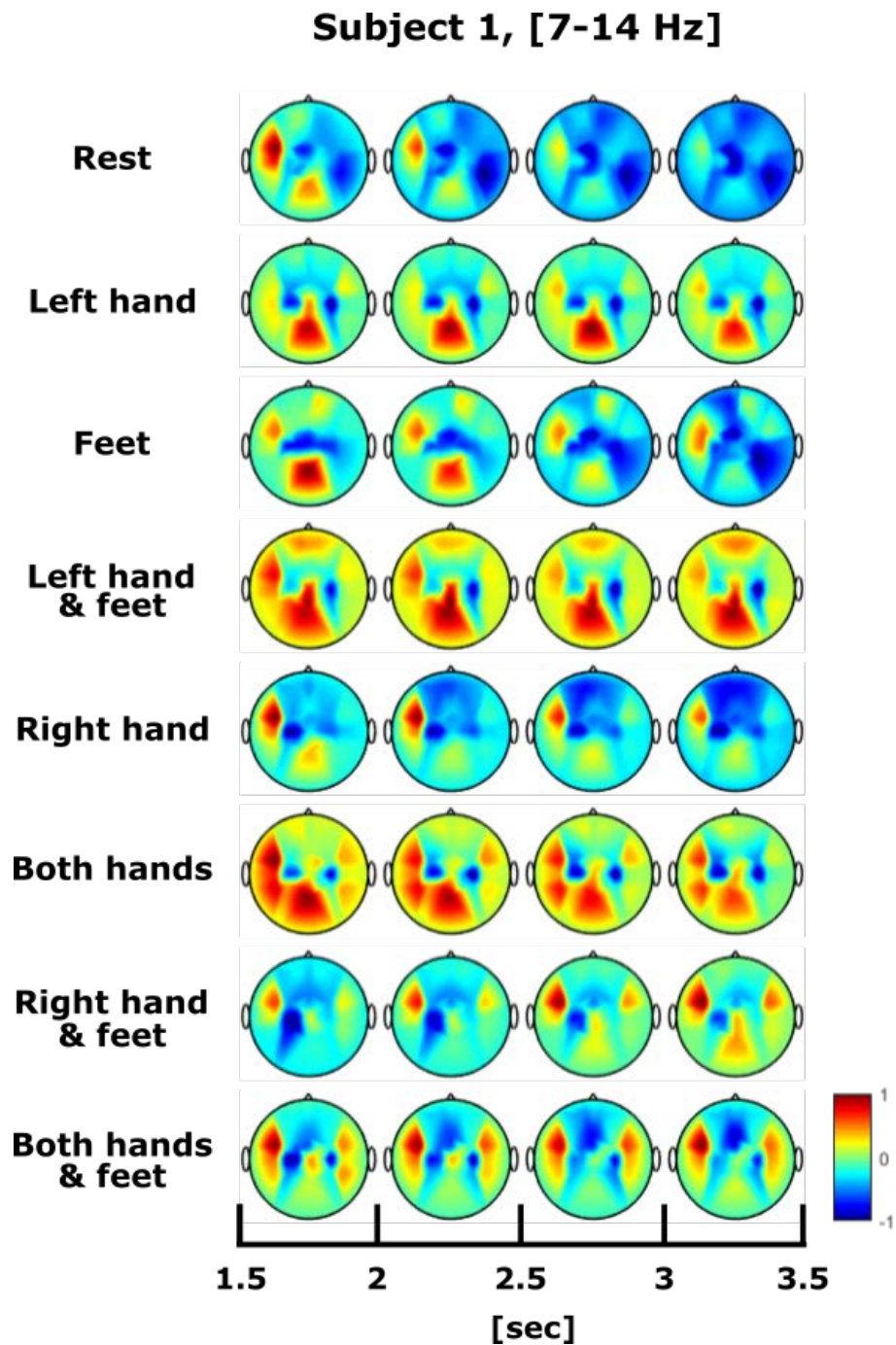
Fig. B.1: Oscillatory power for subject 1 within the alpha band.

Subject 1, [8-30 Hz]
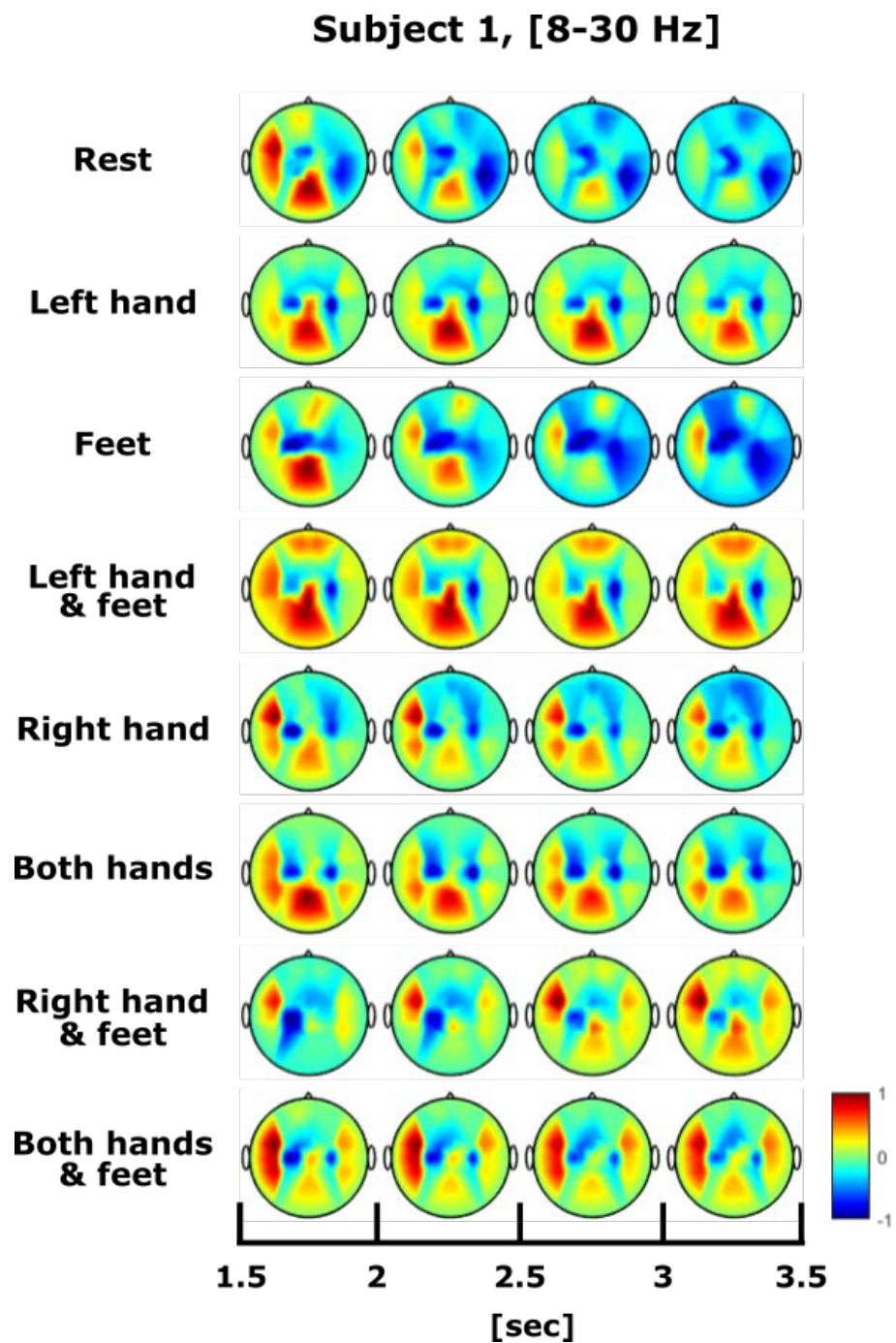
Fig. B.2: Oscillatory power for subject 1 within the alpha and beta bands.
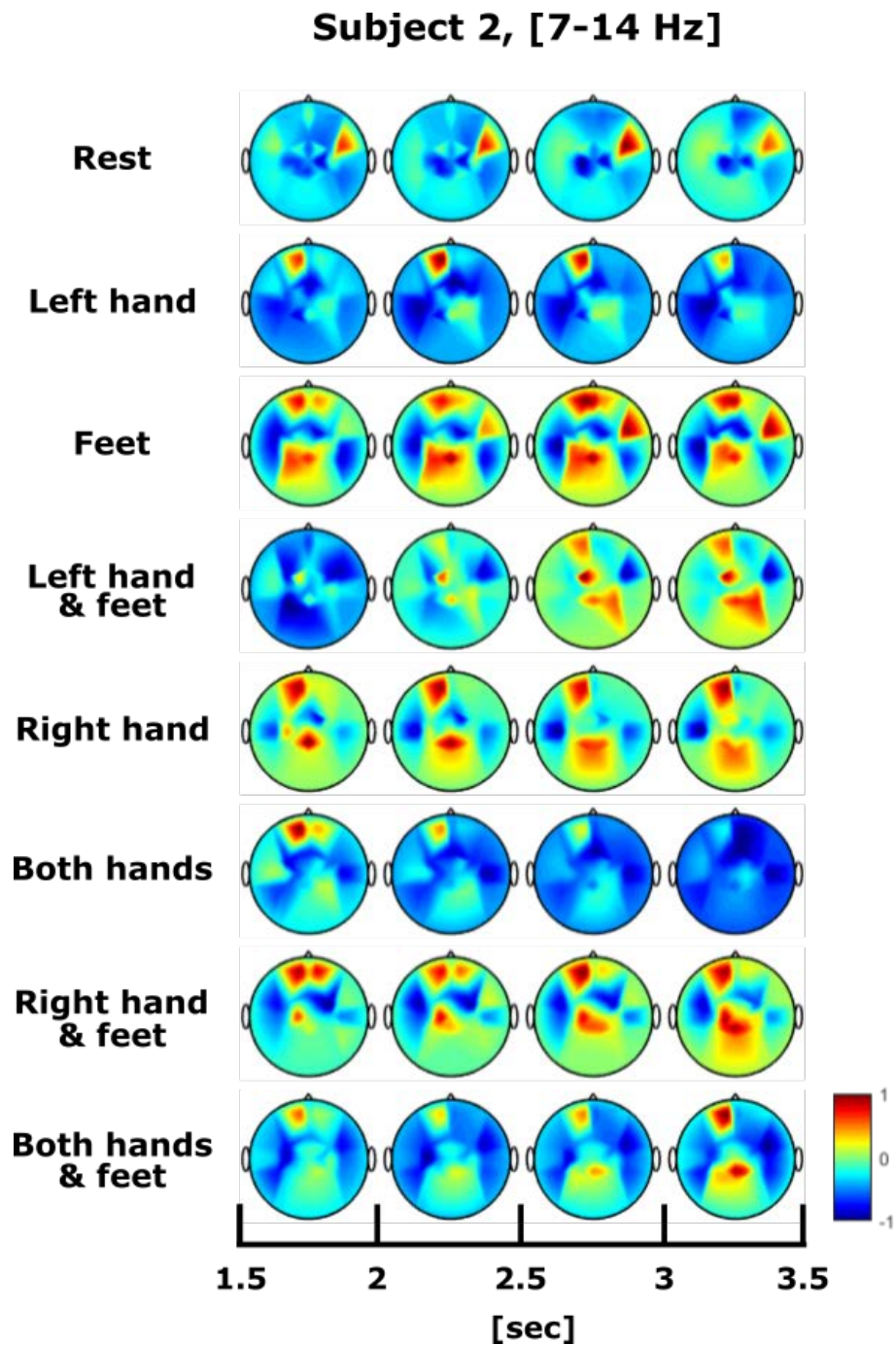
Fig. B.3: Oscillatory power for subject 2 within the alpha band.
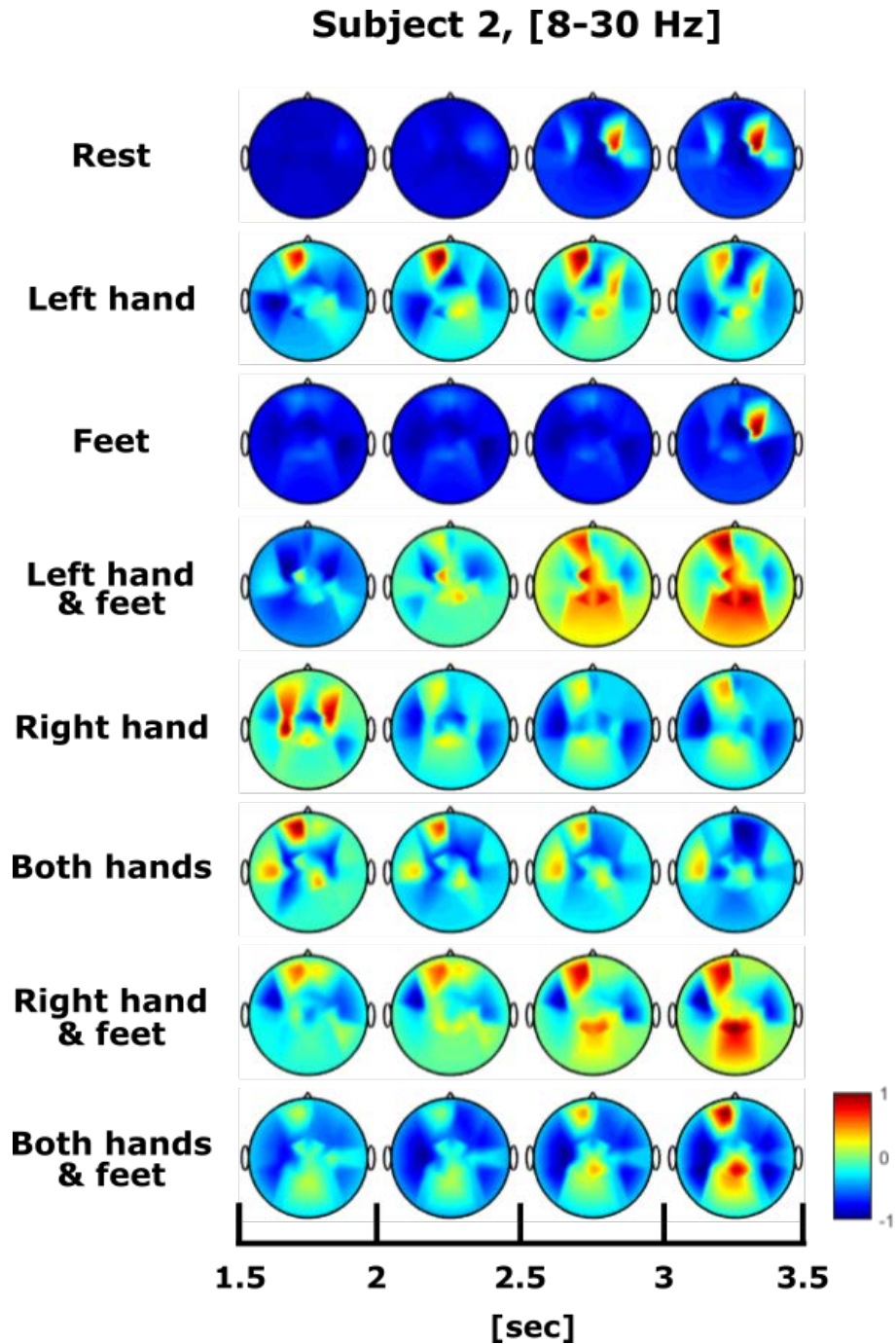
Fig. B.4: Oscillatory power for subject 2 within the alpha and beta bands.
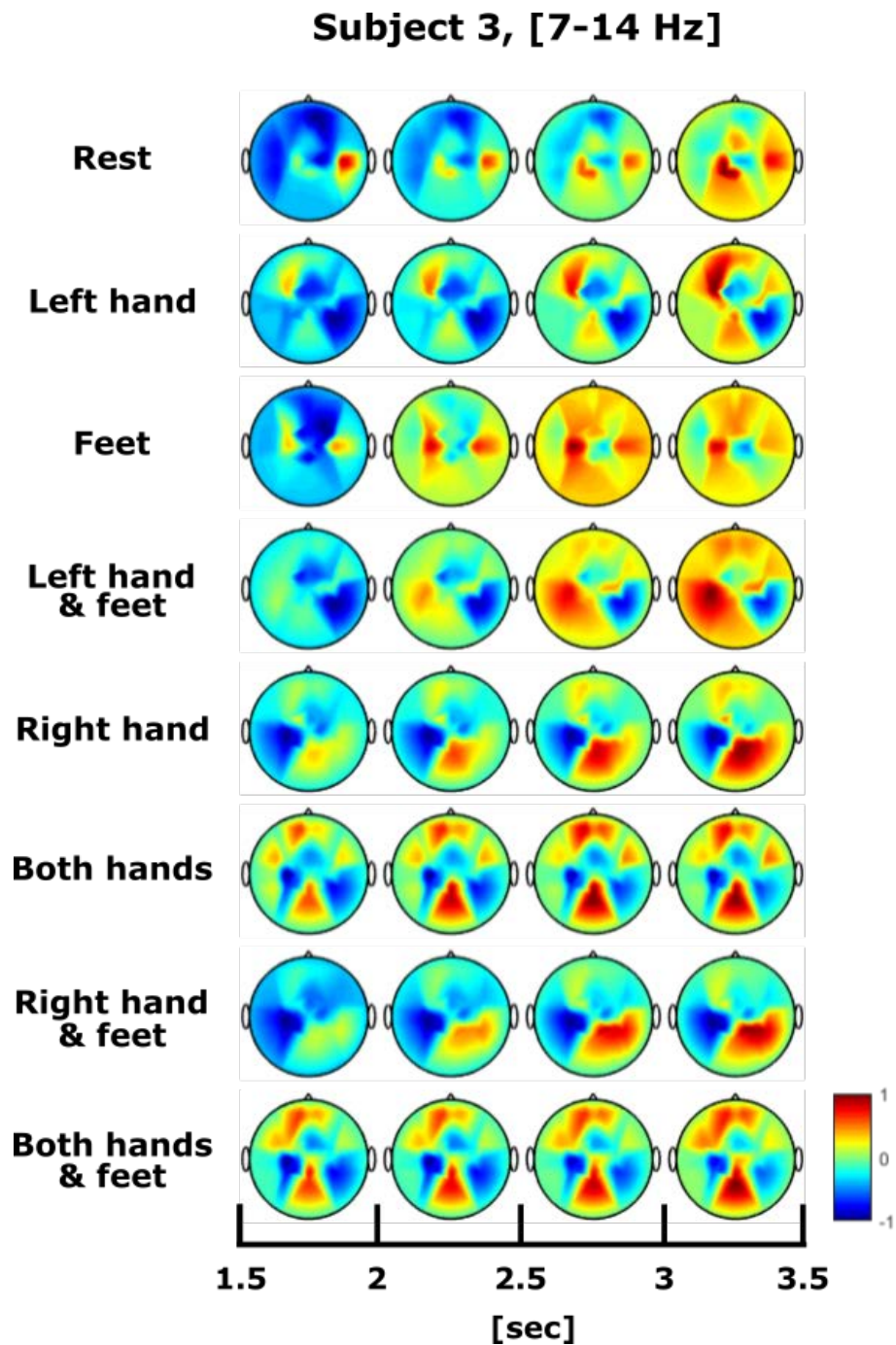
Fig. B.5: Oscillatory power for subject 3 within the alpha band.
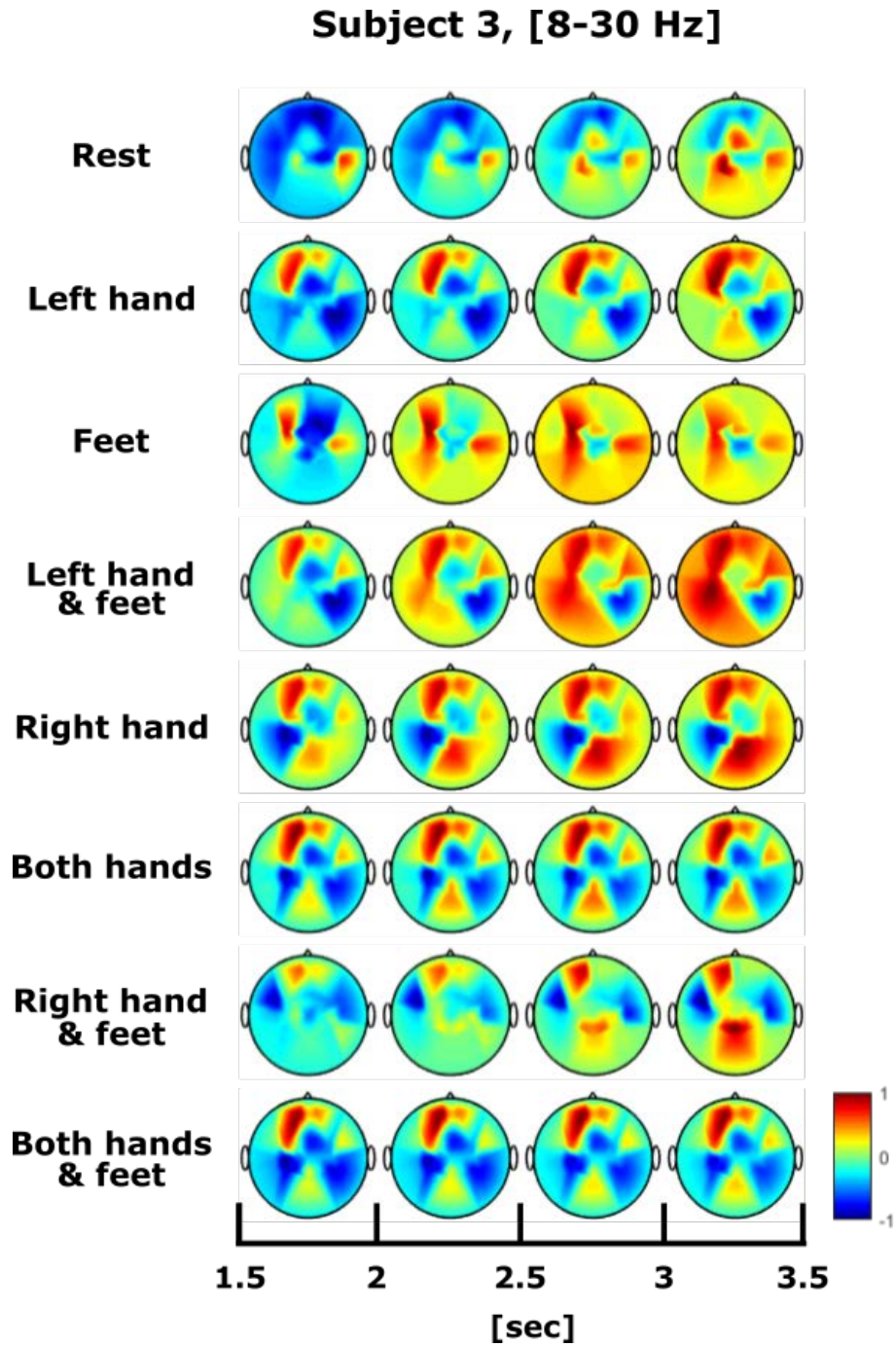
Fig. B.6: Oscillatory power for subject 3 within the alpha and beta bands.
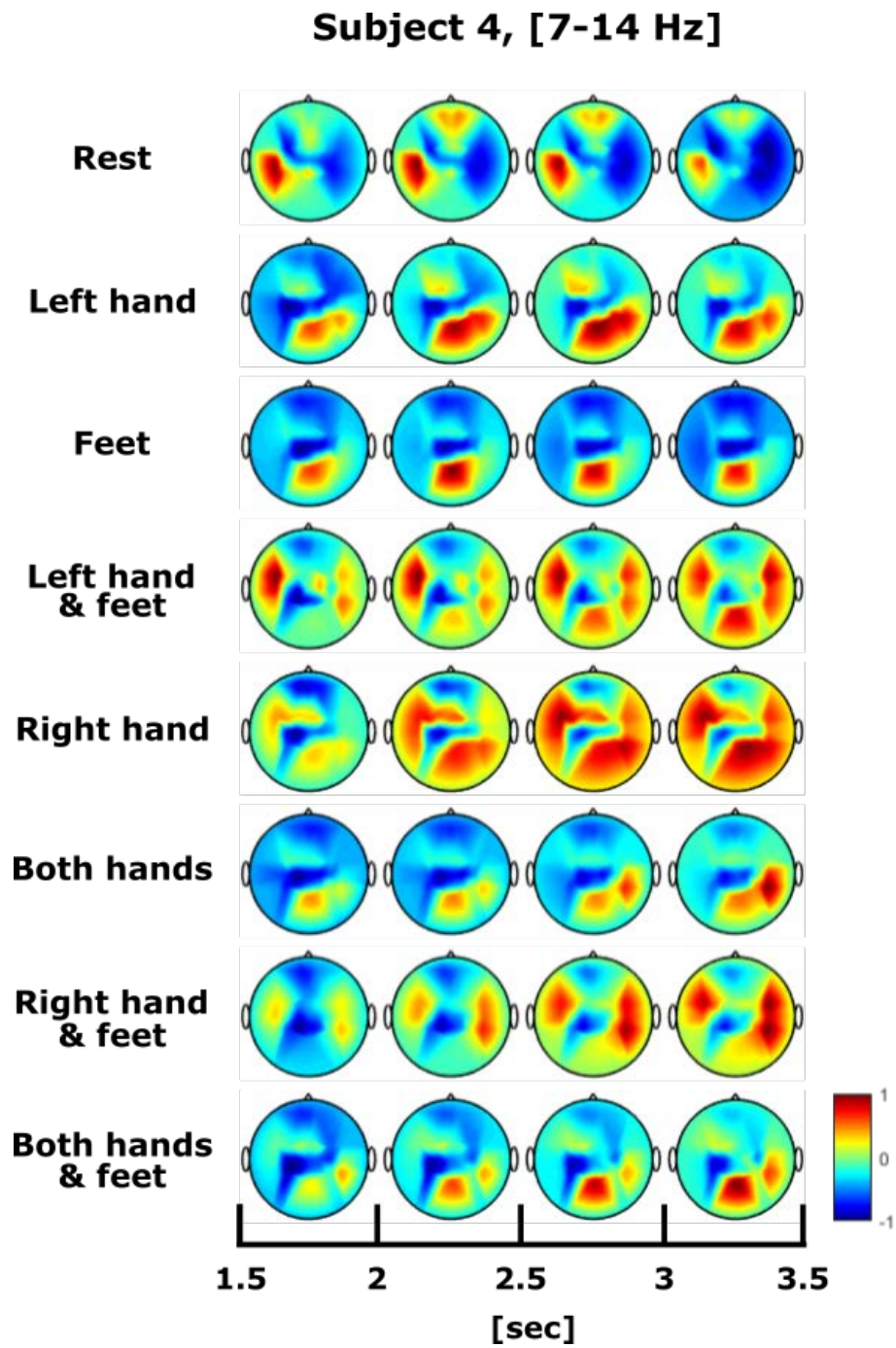
Fig. B.7: Oscillatory power for subject 4 within the alpha band.

Fig. B.8: Oscillatory power for subject 4 within the alpha and beta bands.
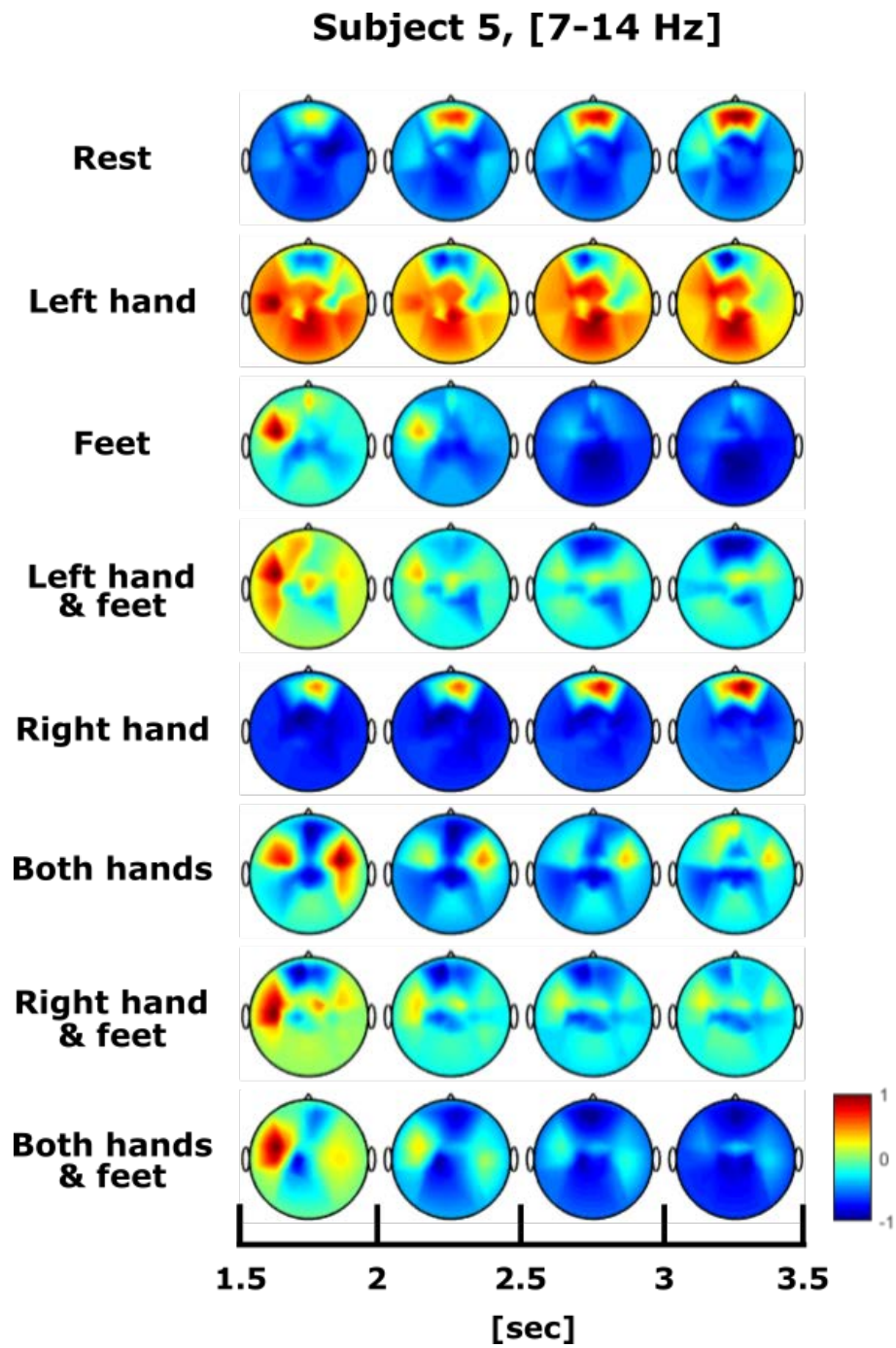
Fig. B.9: Oscillatory power for subject 5 within the alpha band.

Fig. B.10: Oscillatory power for subject 5 within the alpha and beta bands.
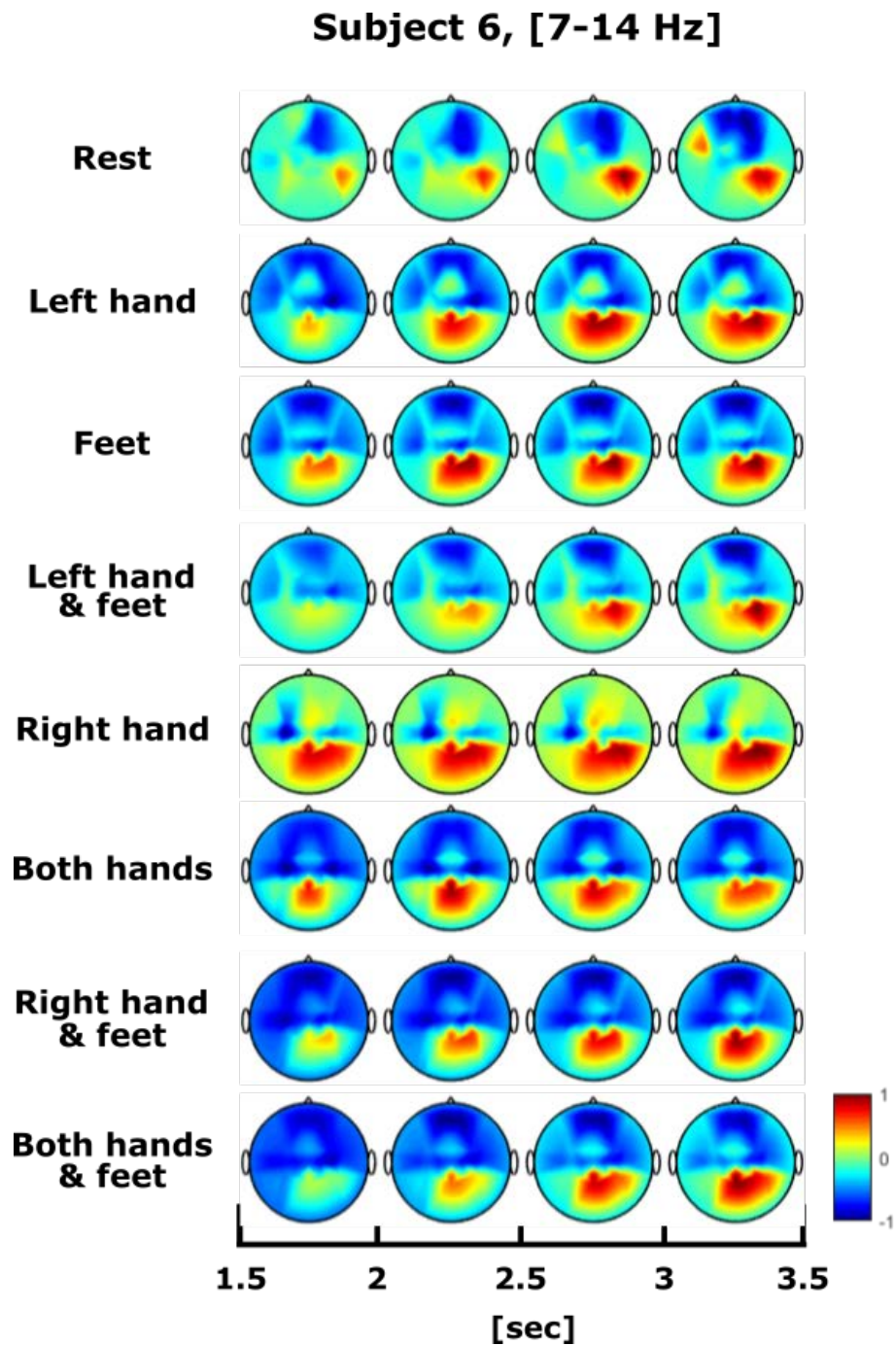
Fig. B.11: Oscillatory power for subject 6 within the alpha band.

Fig. B.12: Oscillatory power for subject 6 within the alpha and beta bands.
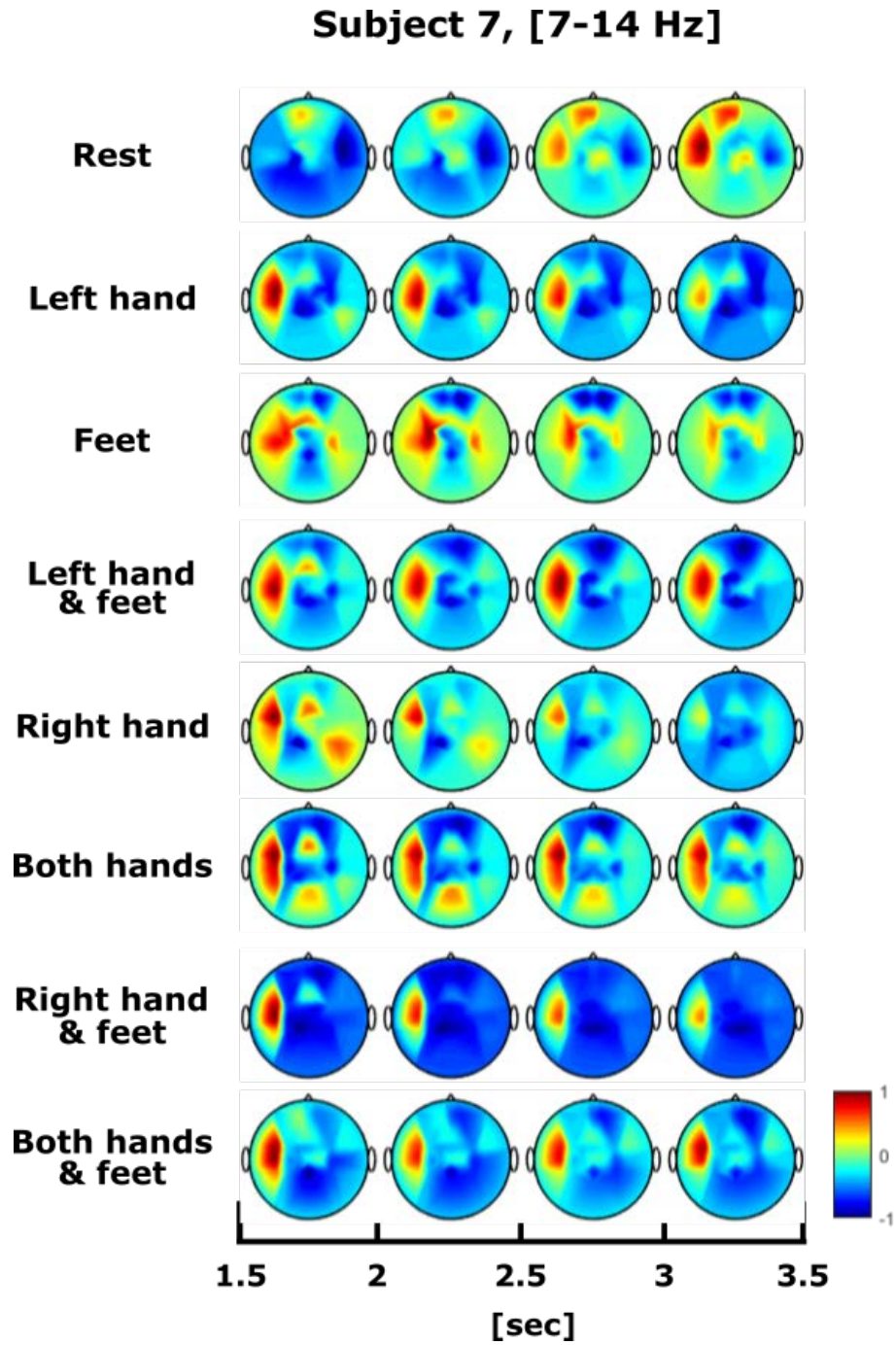
Fig. B.13: Oscillatory power for subject 7 within the alpha band.

Fig. B.14: Oscillatory power for subject 7 within the alpha and beta bands.
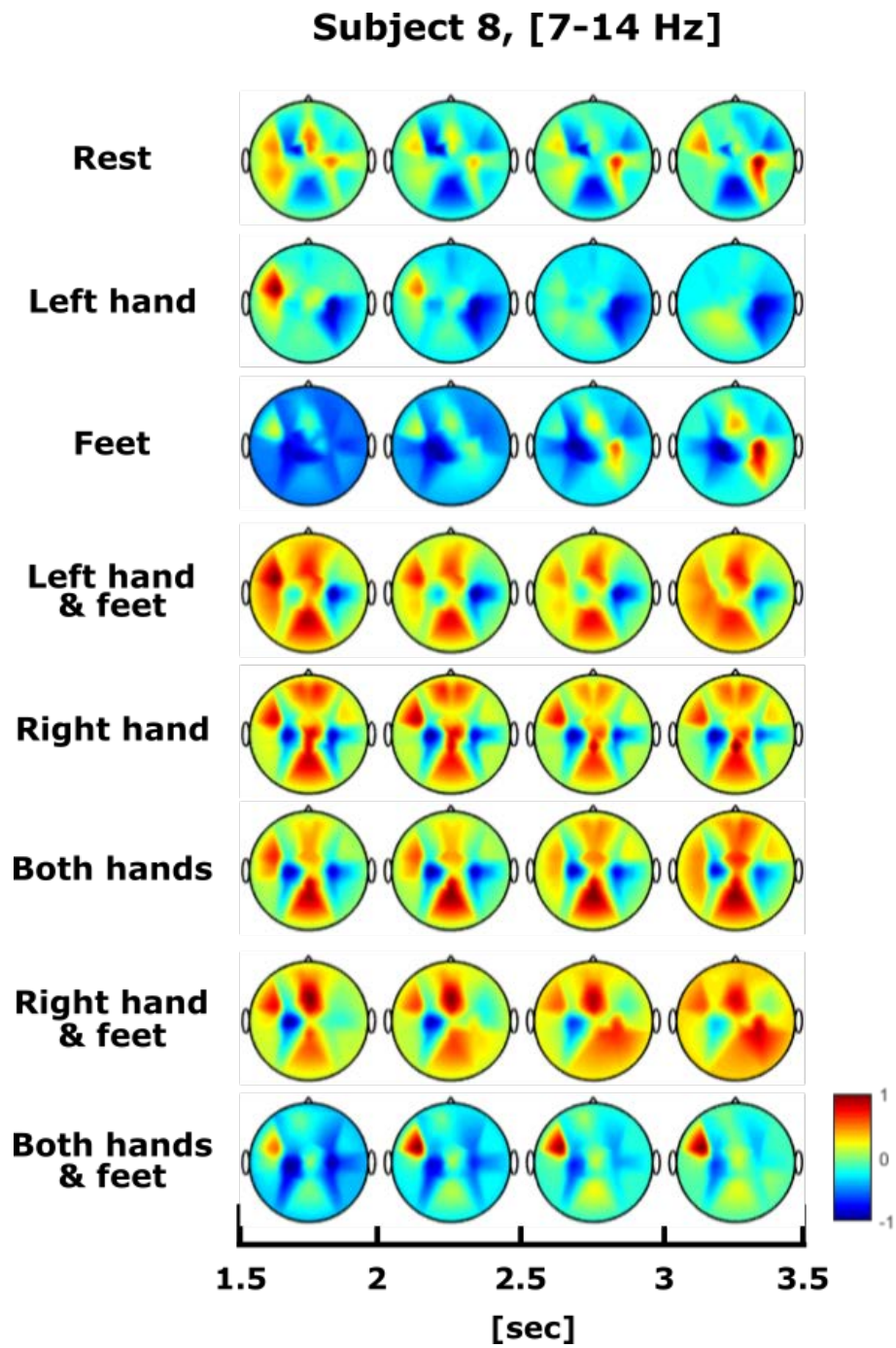
Fig. B.15: Oscillatory power for subject 8 within the alpha band.
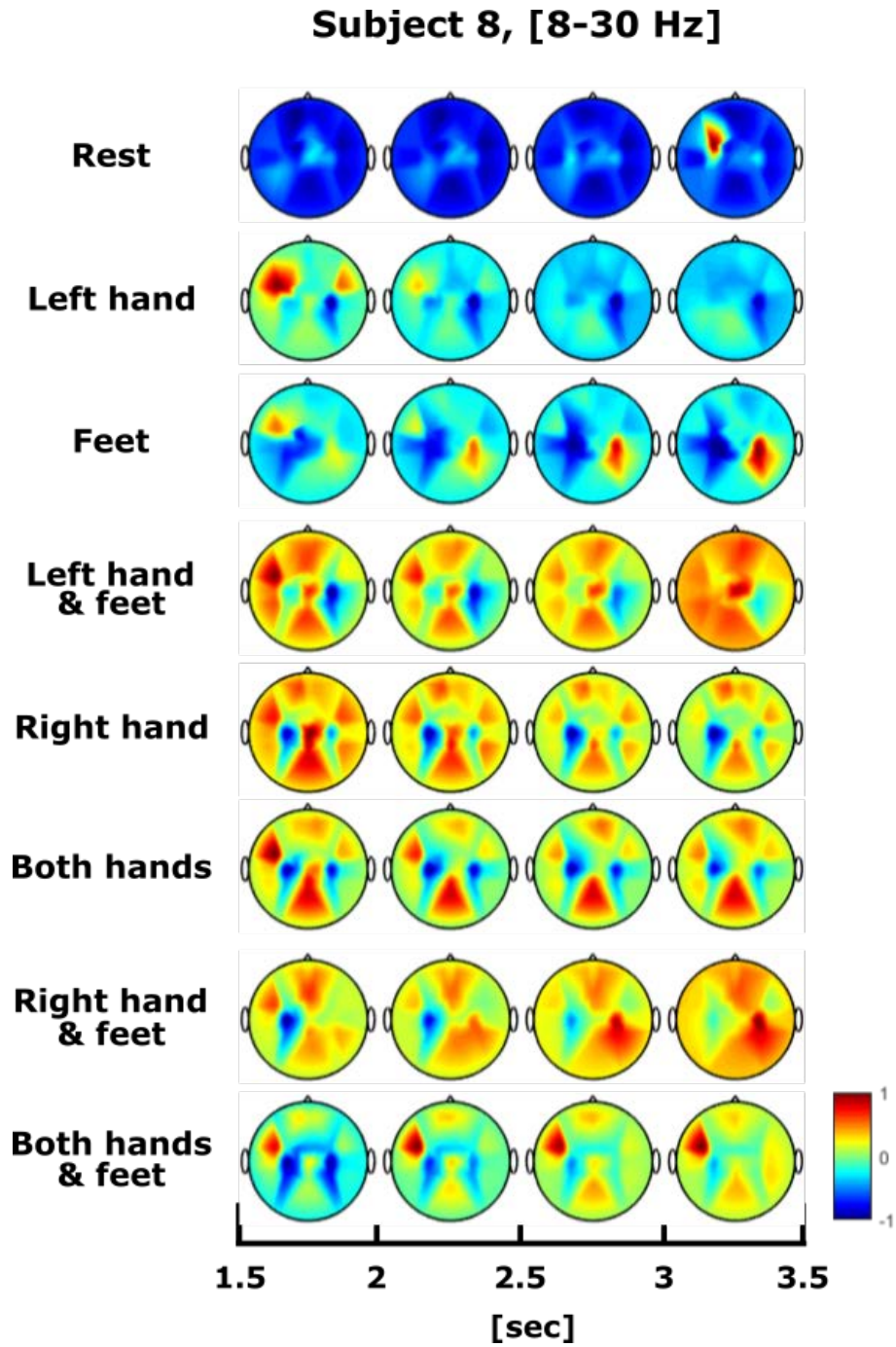
Fig. B.16: Oscillatory power for subject 8 within the alpha and beta bands.
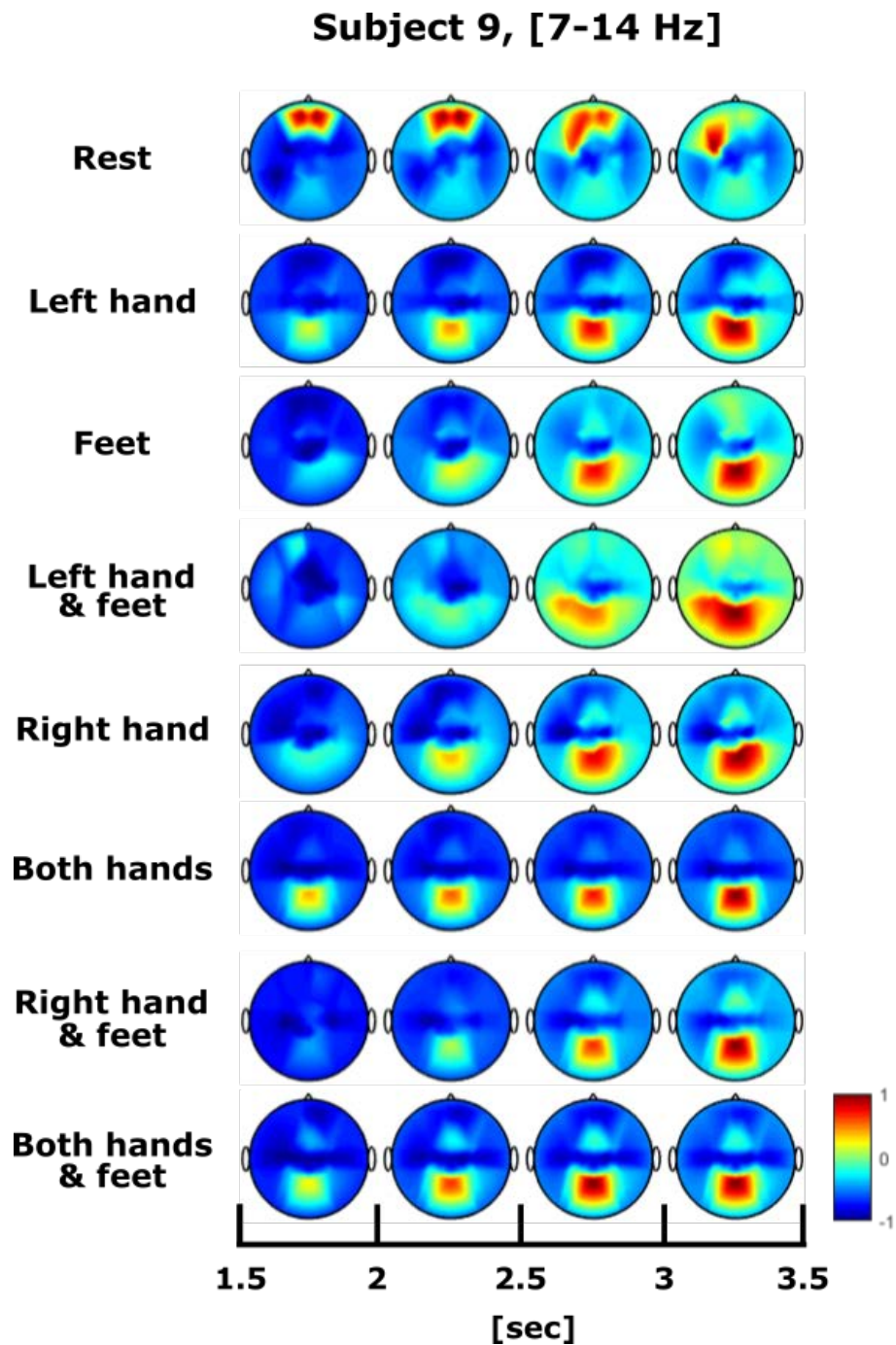
Fig. B.17: Oscillatory power for subject 9 within the alpha band.

Fig. B.18: Oscillatory power for subject 9 within the alpha and beta bands.
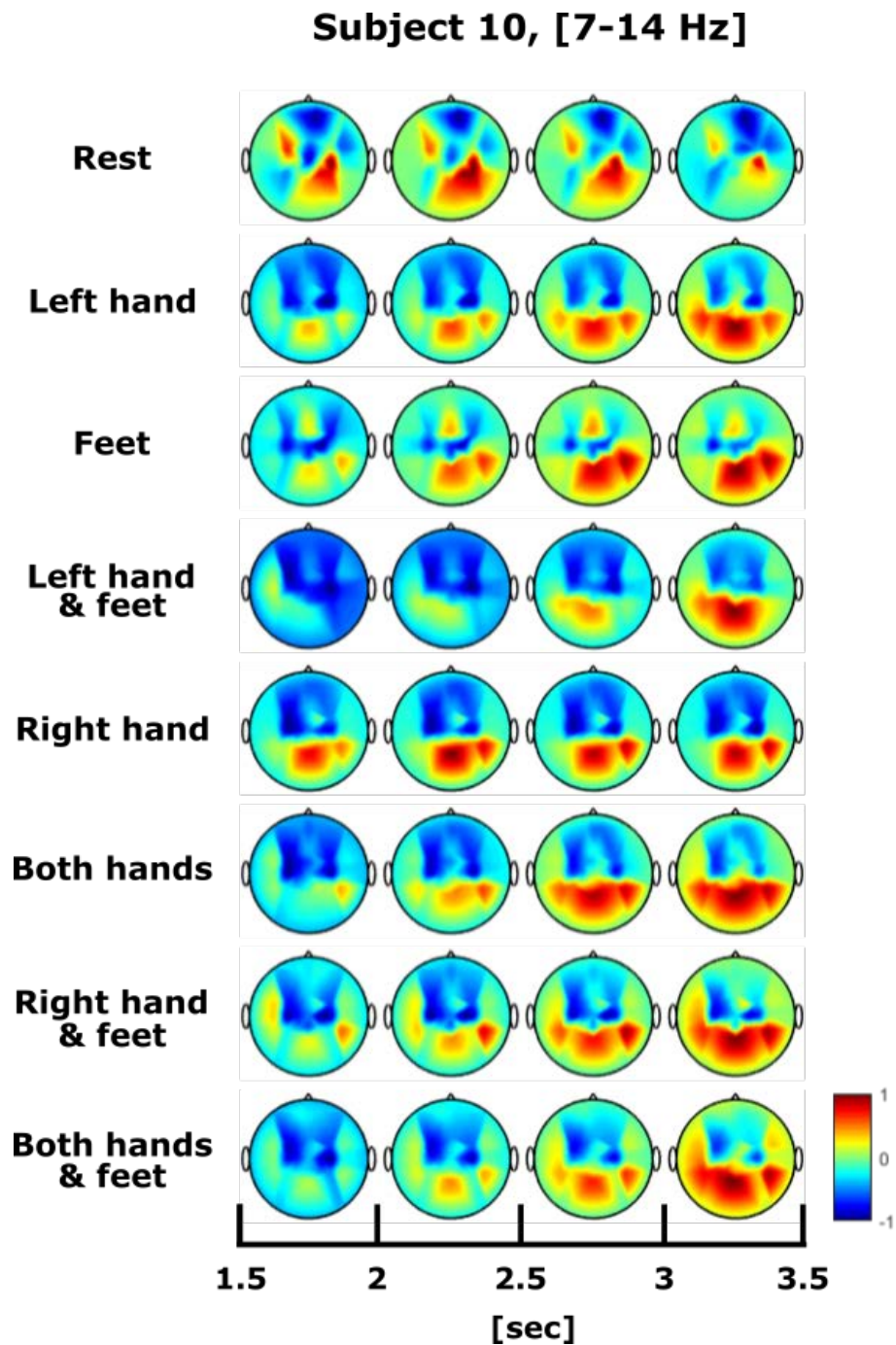
Fig. B.19: Oscillatory power for subject 10 within the alpha band.

Fig. B.20: Oscillatory power for subject 10 within the alpha and beta bands.
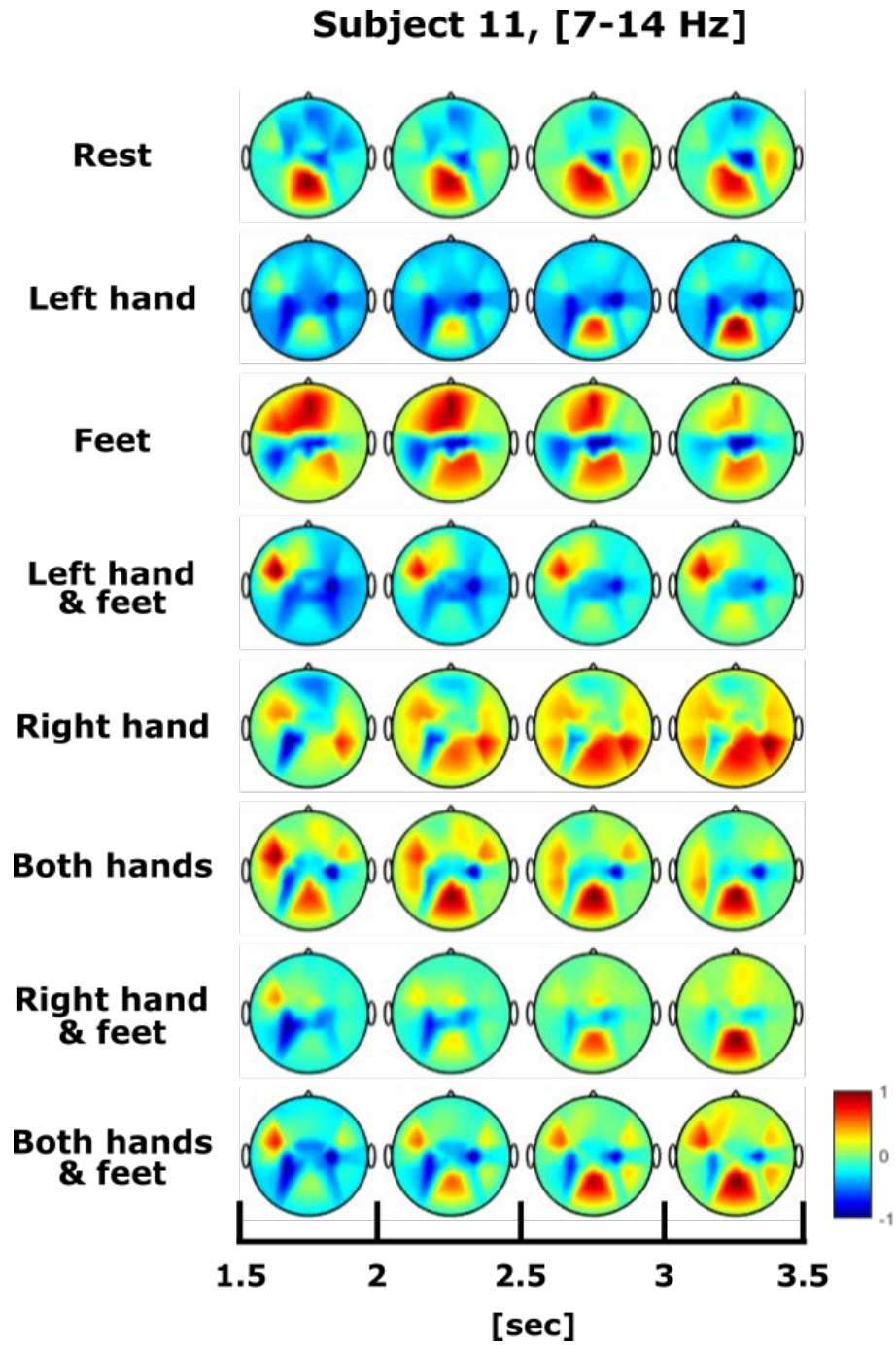
Fig. B.21: Oscillatory power for subject 11 within the alpha band.

Fig. B.22: Oscillatory power for subject 11 within the alpha and beta bands.

C

Offline Results

Along this section the detailed description of the offline results presented in chapter 6 will be presented, which consist of the accuracy yielded by each subject for each one of the described methods. The first case corresponds with the results generated by the standard **CSP** algorithm described in section 4.1.3, together with its analytical version; the **ACSP** approach, which has been discussed in section 4.2. Subsequently, we report the performance obtained by the **FBCSP** described in section 4.4, and by its analytical version; the Filter Bank Analytical Common Spatial Pattern (**FBACSP**) algorithm, which corresponds with one of the contributions of the present work. Afterwards, we present another variant of the CSP method from which feature extraction is directly achieved by considering all classes at once, without having to resort to any complementary technique, namely, the **CSP by JAD** approach (see section 4.5). Since the aforementioned method consists in recursively approximating a transformation matrix for the joint diagonalization of real covariance matrices, its analytical version is not considered within the scope of the present work.

As discussed in section 4.6.2.1, the **MDRM** consists of another approach that allows to directly extract features without using any auxiliary technique. This method has been applied over both databases as an alternative technique for sensorimotor rhythms discrimination. Moreover, we have also proposed a convenient variant of the MDRM method, where the CSP algorithm is applied before computing the covariance matrices that define the Riemannian manifold, which are used as the EEG descriptors, so that more discriminative features are generated (see section 4.6.2.2).

Finally, results obtained by the three contributed algorithms, namely, OsM approach, HM

approach, and OsHM approach, in combination with all the aforementioned feature extraction methods are presented for both databases.

All feature extraction methods (excepting the Riemannian distance-based schema, which do not require a complementary classifier), where applied together with Linear Discriminant Analysis (LDA, see section 4.6.1.3) for classification. In the case of the 4-*class* database, results by using Support Vectors Machines (SVM, see section 4.6.1.4)are also presented.

C.1 Results: CSP method

In this section we present the offline results generated by applying the standard CSP algorithm discussed in section 4.1.3 over both databases and for all subjects. As mentioned, the number of the m selected pairs was set to $m=2$, which generates a total of four spatial filters for the corresponding feature extraction.

Accuracy performance by using the discussed classification algorithms under the one-versus-one approach is reported on tables C.1 and C.3 for the 4-*class* and 8-*class* databases respectively. Moreover, tables C.2 and C.4 show, respectively, the yielded accuracy under the one-versus-all approach over the 4-*class* and 8-*class* databases.

Table C.1: Accuracy results for the CSP method using one-versus-one classification on the 4-class database

| | **Classification method** | | | | |
| | | | **SVM** | | |
Subject	**LDA**	**QDA**	**Linear**	**Polynomial**	**RBF**
1	50.94±8.06%	51.88±8.07%	50.00±2.70%	40.63±4.62%	47.50±7.29%
2	41.88±5.54%	38.44±4.72%	42.81±8.62%	37.81±2.58%	39.06±4.72%
3	61.25±24.37%	60.31±24.01%	54.06±19.62%	48.75±22.01%	47.81±17.45%
4	60.00±5.95%	51.56±3.13%	57.19±3.87%	49.69±13.75%	50.94±5.94%
5	53.44±4.13%	50.94±3.76%	52.81±4.72%	51.56±6.49%	50.62±9.82%
AVG	53.50±9.26%	51.75±8.88%	50.56±8.06%	45.69±9.89%	47.19±8.90%

Table C.2: Accuracy results for the CSP method using one-versus-all classification on the 4-class database

| | **Classification method** | | | | |
| | | | **SVM** | | |
Subject	**LDA**	**QDA**	**Linear**	**Polynomial**	**RBF**

1	49.69±6.32%	46.25±5.68%	42.50±4.45%	41.87±4.27%	31.56±8.00%
2	42.81±6.40%	44.69±5.98%	42.19±6.07%	36.88±8.07%	30.63±5.05%
3	59.69±23.99%	59.06±22.37%	37.19±5.90%	40.00±10.80%	32.19±8.19%
4	60.31±5.72%	59.38±3.31%	45.31±5.90%	37.81±6.16%	30.63±7.25%
5	56.87±5.15%	52.81±6.07%	41.56±1.20%	38.75±1.77%	28.75±1.77%
AVG	53.88±9.52%	52.44±8.68%	41.75±4.70%	39.06±6.21%	30.75±1.77%

Table C.3: Accuracy results for the CSP method using one-versus-one classification on the 8-class database

	Classification method	
Subject	**LDA**	**QDA**
1	36.25±16.20%	33.75±14.74%
2	47.50±12.88%	45.83±14.06%
3	79.17±15.16%	78.75±15.10%
4	48.75±17.17%	47.92±18.59%
5	23.33±14.58%	22.50±15.54%
6	37.50±15.04%	37.50±13.53%
7	35.00±13.29%	38.33±15.37%
8	50.00±19.42%	44.58±17.88%
9	33.75±17.10%	32.08±14.56%
10	61.25±19.24%	62.08±17.52%
11	53.75±16.13%	52.08±15.77%
AVG	46.02±16.02%	45.04±15.70%

Table C.4: Accuracy results for the CSP method using one-versus-all classification on the 8-class database

	Classification method	
Subject	**LDA**	**QDA**
1	31.67±13.43%	33.33±14.06%
2	45.42±14.11%	45.00±15.26%
3	71.25±15.79%	70.00±16.28%
4	46.67±17.96%	43.33±14.21%
5	25.00±14.31%	22.50±14.46%

6	36.67±14.66%	35.42±16.77%
7	36.67±15.37%	34.58±15.29%
8	38.75±18.38%	35.42±16.77%
9	27.08±13.16%	25.42±16.24%
10	53.75±18.61%	55.83±17.90%
11	54.58±14.49%	57.08±18.18%
AVG	42.50±15.48%	41.63±15.95%

C.2 Results: ACSP method

Tables C.5 and C.7 present, respectively, the offline results generated over the 4-*class* and 8-*class* databases for all subjects by the ACSP method under the one-versus-one approach. In addition, the yielded performance under the one-versus-all approach is presented in table C.6 for the 4-*class* database, and in table C.8, for the 8-*class* one. As with the standard CSP method, we have decided to use $m=2$ pairs of spatial filters (4 in total) to extract the features generated by its analytical version described in section 4.2. Note that, since the ACSP method generates complex features, it is necessary to use a special strategy for applying the SVM approach (which is out of the scope of the present work); therefore for the 4-*class* database, we only present the results generated in combination with Linear Discriminant Analysis classification.

Table C.5: Accuracy results for the ACSP method using one-versus-one classification on the 4-class database

	Classification method	
Subject	**LDA**	**QDA**
1	51.25±5.68%	52.19±5.14%
2	42.19±3.59%	37.50±8.84%
3	61.25±24.30%	59.06±22.72%
4	57.50±3.95%	56.25±3.95%
5	56.25±6.45%	55.31±5.44%
AVG	53.69±7.88%	52.06±9.22%

Table C.6: Accuracy results for the ACSP method using one-versus-all classification on the 4-class database

Subject	Classification method	
	LDA	**QDA**
1	51.25±6.21%	51.56±8.68%
2	39.06±6.80%	40.31±2.58%
3	65.31±18.69%	62.19±24.84%
4	58.75±5.68%	57.81±7.80%
5	57.19±5.14%	56.88±4.84%
AVG	54.31±8.50%	53.75±9.75%

Table C.7: Accuracy results for the ACSP method using one-versus-one classification on the 8-class database

Subject	Classification method	
	LDA	**QDA**
1	32.50±15.61%	35.00±15.88%
2	49.58±12.92%	46.25±17.72%
3	80.42±12.14%	31.67±9.13%
4	54.58±19.28%	54.17±17.47%
5	23.33±14.95%	22.08±13.80%
6	43.33±17.90%	41.67±17.47%
7	37.08±14.89%	40.00±16.54%
8	47.08±19.33%	40.42±16.96%
9	33.75±18.02%	34.58±17.27%
10	60.42±17.08%	52.50±14.83%
11	55.42±18.77%	53.75±18.32%
AVG	47.04±16.44%	41.10±15.94%

Table C.8: Accuracy results for the ACSP method using one-versus-all classification on the 8-class database

Subject	Classification method	
	LDA	**QDA**
1	32.08±15.29%	29.58±14.86%

2	47.08±16.96%	49.58±13.73%
3	71.25±15.45%	70.42±14.11%
4	49.17±19.12%	43.33±15.99%
5	29.17±13.67%	26.25±13.27%
6	38.33±15.02%	37.08±15.22%
7	32.92±18.42%	39.58±15.77%
8	40.00±16.87%	39.58±15.43%
9	28.33±18.26%	27.92±15.63%
10	57.08±19.33%	60.00±16.54%
11	53.75±15.10%	60.00±15.88%
AVG	43.56±16.68%	43.94±15.13%

C.3   Results: FBCSP method

This section presents the accuracy offline results generated with the FBCSP method. For its implementation, as it was discussed in section 4.4, we have decided to use $m=2$ spatial filters (4 in total) to extract the CSP features within each one of the nine frequency bands considered by the filter bank. This process generates feature vectors of 36 elements each, from which only the 4 most discriminative components are selected by using the Mutual Information-based Best Individual Feature (MIBIF) algorithm. Note that, since the selected features are paired, this number fluctuates between 4 components (if the selected elements come from the same 2 pairs) to 8 (if each one of them belongs to a different pair). Tables C.9 and C.11 show the accuracy yielded by the FBCSP method applied in combination with the considered classification algorithms under a one-versus-one approach for the 4-*class* and 8-*class* databases respectively. Moreover, the performance obtained under the one-versus-all approach is presented in table C.10 for the 4-*class* database, and in table C.12, for the 8-*class* one.

Table C.9: Accuracy results for the FBCSP method using one-versus-one classification on the 4-class database

			Classification method		
				SVM	
Subject	**LDA**	**QDA**	**Linear**	**Polynomial**	**RBF**
1	48.12±8.32%	46.25±6.21%	44.06±3.87%	48.75±5.86%	36.25±4.46%
2	35.00±9.30%	32.19±8.00%	38.44±3.44%	34.69±5.53%	34.37±6.50%
3	69.37±14.31%	68.12±14.01%	58.75±14.58%	50.63±17.75%	40.31±14.66%
4	53.12±6.57%	52.81±3.29%	47.81±xx.xx%	53.44±xx.xx%	47.81±xx.xx%

5	52.19±5.34%	47.19±8.06%	51.87±5.45%	52.50±3.68%	46.88±5.15%
AVG	51.56±8.77%	49.31±7.91%	48.19±7.44%	48.00±7.28%	41.13±7.52%

Table C.10: Accuracy results for the FBCSP method using one-versus-all classification on the 4-class database

	Classification method				
				SVM	
Subject	**LDA**	**QDA**	**Linear**	**Polynomial**	**RBF**
1	45.62±6.65%	47.19±8.00%	41.25±3.23%	41.56±3.29%	34.38±9.44%
2	38.44±8.00%	38.75±3.54%	36.25±3.68%	34.38±3.89%	29.38±9.44%
3	70.31±10.38%	63.44±18.99%	63.12±8.93%	62.81±11.24%	43.75±19.55%
4	52.19±4.13%	54.37±3.89%	53.75±2.70%	52.19±3.13%	44.69±6.40%
5	54.69±7.17%	53.75±5.10%	56.56±7.17%	48.13±6.81%	39.06±6.24%
AVG	52.25±7.27%	51.50±7.90%	50.19±5.14%	47.81±5.67%	38.25±9.20%

Table C.11: Accuracy results for the FBCSP method using one-versus-one classification on the 8-class database

	Classification method	
Subject	**LDA**	**QDA**
1	27.92±14.93%	29.17±16.19%
2	41.25±15.45%	41.25±15.79%
3	89.17±9.13%	89.17±8.42%
4	50.42±20.36%	45.83±17.47%
5	23.75±15.17%	22.08±14.93%
6	46.25±16.79%	45.83±18.08%
7	40.42±16.96%	40.00±14.46%
8	48.33±19.07%	45.83±20.06%
9	43.33±16.65%	38.33±18.26%
10	49.17±15.72%	50.00±17.37%
11	58.33±15.51%	56.25±16.33%
AVG	47.12±15.98%	45.79±16.12%

Table C.12: Accuracy results for the FBCSP method using one-versus-all classification on the 8-class database

	Classification method	
Subject	**LDA**	**QDA**
1	31.25±14.21%	25.00±13.13%
2	28.75±17.42%	32.50±17.25%
3	65.00±13.69%	66.25±17.42%
4	42.08±16.24%	39.17±16.65%
5	24.17±12.69%	17.50±12.97%
6	38.75±18.96%	37.92±16.57%
7	37.50±17.68%	36.67±19.68%
8	47.08±15.63%	44.58±14.56%
9	37.08±14.11%	32.92±15.57%
10	44.58±19.05%	43.75±15.31%
11	48.33±17.29%	49.58±16.24%
AVG	40.42±16.27%	38.71±15.94%

C.4   Results: FBACSP method

The analytical version of the FBCSP method discussed in section 4.4 has been also implemented as another contribution of the present work, to which we will refer as the Filter Bank Analytical Common Spatial Pattern (FBACSP) algorithm. For its implementation, the same parameters of the standard FBCSP method have been selected, the main difference is that feature extraction is carried out according to the ASCP algorithm discussed in section 4.2. In this way, both feature extraction and classification, were carried out by following a one-versus-one approach (see tables C.13 and C.15 for the 4-*class* and 8-*class* databases respectively), and under a one-versus-all approach (see tables C.14 for the 4-*class* database and C.16 for the 8-*class* one). As mentioned, given that the ACSP method generates complex features, it is necessary to use a special strategy for applying the SVM approach; therefore, we only present the results generated in combination with Linear Discriminant Analysis classification.

Table C.13: Accuracy results for the FBACSP method using one-versus-one classification on the 4-class database

	Classification method	
Subject	**LDA**	**QDA**
1	51.25±3.68%	48.44±5.14%

2	39.69±5.81%	38.44±6.64%
3	60.63±22.86%	62.19±20.14%
4	45.94±2.77%	48.44±2.58%
5	52.81±6.80%	55.00±8.10%
AVG	50.06±8.38%	50.50±8.52%

Table C.14: Accuracy results for the FBACSP method using one-versus-all classification on the 4-class database

	Classification method	
Subject	**LDA**	**QDA**
1	47.19±4.49%	48.13±6.17%
2	43.12±8.75%	41.56±5.14%
3	61.25±21.48%	59.69±18.38%
4	45.00±4.21%	44.06±6.80%
5	47.81±6.95%	46.56±5.04%
AVG	48.88±9.18%	48.00±8.31%

Table C.15: Accuracy results for the FBACSP method using one-versus-one classification on the 8-class database

	Classification method	
Subject	**LDA**	**QDA**
1	27.08±13.16%	25.42±12.49%
2	35.00±17.49%	36.25±17.17%
3	86.67±12.25%	26.67±13.43%
4	42.50±14.16%	39.58±14.34%
5	25.00±13.93%	26.67±14.58%
6	49.17±16.06%	47.50±14.08%
7	35.00±13.69%	37.08±13.73%
8	44.58±15.63%	39.17±16.97%
9	39.17±18.49%	37.08±18.42%
10	52.92±16.96%	49.58±19.28%
11	50.00±16.08%	49.17±16.06%
AVG	44.28±15.26%	37.65±15.5%

Table C.16: Accuracy results for the FBACSP method using one-versus-all classification on the 8-class database

	Classification method	
Subject	**LDA**	**QDA**
1	24.17±18.84%	20.00±15.95%
2	30.83±16.33%	32.92±12.92%
3	69.17±13.43%	68.75±13.03%
4	33.33±16.52%	36.67±15.02%
5	23.75±10.56%	24.58±13.33%
6	40.00±14.83%	36.25±11.54%
7	29.17±14.80%	28.75±16.46%
8	32.08±16.64%	34.17±13.51%
9	28.75±15.10%	34.58±15.63%
10	44.17±11.24%	38.33±13.51%
11	49.58±19.28%	53.33±15.37%
AVG	36.82±15.23%	37.12±14.21%

C.5   Results: CSP by JAD method

In this section we present the results generated after applying the CSP by JAD algorithms. As mentioned in section 4.5, feature extraction under this method is directly achieved by considering all classes at once, without having to resort to any complementary technique, such as the one-versus-one or one-versus-all approaches. Tables C.17 and C.18 show the accuracy results obtained after applying the algorithm over the the 4-*class* and 8-*class* databases respectively.

Table C.17: Accuracy results for the CSP by JAD method on the 4-class database

	Classification method				
			SVM		
Subject	**LDA**	**QDA**	**Linear**	**Polynomial**	**RBF**
1	49.38±7.18%	45.31±4.83%	48.75±9.52%	36.87±3.75%	48.12±5.15%
2	38.44±2.77%	35.94±5.14%	38.12±6.81%	34.06±3.29%	35.00±4.89%
3	60.00±24.60%	58.75±24.26%	46.56±17.72%	48.75±17.23%	38.44±11.47%
4	45.31±16.84%	38.75±13.81%	38.12±10.43%	36.88±9.71%	37.81±11.88%
5	55.63±3.89%	48.44±7.10%	52.50±5.20%	48.13±8.75%	49.75±5.50%
AVG	49.75±11.06%	45.44±11.03%	44.81±9.94%	40.94±8.55%	41.62±7.78%

Table C.18: Accuracy results for the CSP by JAD method using on the 8-class database

	Classification method	
Subject	**LDA**	**QDA**
1	22.08±13.00%	20.42±12.92%
2	53.33±13.51%	42.50±20.92%
3	82.50±10.69%	73.75±15.17%
4	42.92±13.00%	35.42±14.34%
5	24.58±12.05%	22.92±15.77%
6	43.33±15.99%	39.17±19.62%
7	33.75±11.90%	29.17±16.19%
8	47.50±17.18%	35.42±14.34%
9	22.92±13.16%	20.00±14.90%
10	62.50±18.28%	55.83±20.16%
11	58.33±15.16%	53.75±18.02%
AVG	44.89±13.99%	38.94±16.58%

C.6   Results: MDRM algorithm

As discussed in section 4.6.2.1, The Minimum Distance to Riemannian Mean (MDRM) algorithm consists of an alternative approach from which features are directly extracted by considering all classes at once, without having to resort to any grouping technique, such as the one-versus-one or one-versus-all approaches. These features consist of the Riemannian distance between the input vectors covariance matrices and the Riemannian class means of the considered conditions (see section 4.6.2.1). Since the predicted labels are simply assigned according to the shortest distance separating the input elements from the class means generated during the training process, it is not necessary either to apply any complementary classification function. Because of this reason, the MDRM results a convenient approach that allows to easily carry out classification.

Tables C.19 and C.20 present, respectively, the accuracy offline results generated by the MDRM algorithm over the 4-*class* and 8-*class* databases.

Table C.19: Accuracy results for the MDRM method on the 4-class database

Subject	Accuracy
1	48.75±8.48%
2	35.00±1.77%

3	58.75±10.75%
4	40.62±2.92%
5	43.75±8.35%
AVG	45.37±6.47%

Table C.20: Accuracy results for the MDRM method on the 8-class database

Subject	Accuracy
1	29.17±12.85%
2	38.75±11.99%
3	71.67±15.72%
4	45.00±16.28%
5	27.92±11.69%
6	15.42±7.10%
7	39.17±15.99%
8	45.83±16.52%
9	20.00±11.18%
10	54.58±14.86%
11	55.83±16.65%
AVG	40.3±13.71%

C.7   Results: CSP + MDRM algorithm

As described in section 4.6.2.2, we have proposed a new classification approach that consists of the combination of the CSP method and the Minimum Distance to Riemannian Mean algorithm. The idea is to generate more discriminative covariance matrices, which are used as descriptors within the Riemannian manifold, by computing them over the CSP projected trials instead of over the band-pass filtered segments, as it is done by the standard MDRM algorithm. Other than that, the method follows the same principle. For generating the CSP filters we have applied both the one-versus-one and the one-versus-all grouping techniques.

Tables C.21 and C.22 show the accuracy results obtained over the 4-*class* and the 8-*class* databases respectively.

Table C.21: Accuracy results for the CSP algorithm applied in combination with the MDRM approach on the 4-class database

Classification method

Subject	OVO	OVA
1	47.19±8.38%	53.12±5.99%
2	34.06±4.00%	42.19±4.83%
3	60.31±21.78%	60.31±23.81%
4	50.62±2.60%	55.00±4.68%
5	47.81±5.72%	55.62±1.61%
AVG	48.00±8.50%	53.25±8.19%

Table C.22: Accuracy results for the CSP algorithm applied in combination with the MDRM approach on the 8-class database

	Classification method	
Subject	**OVO**	**OVA**
1	35.42±16.44%	32.92±16.89%
2	43.75±13.03%	43.33±10.75%
3	76.67±15.30%	67.92±12.58%
4	40.83±11.34%	38.33±15.37%
5	27.50±12.02%	24.58±12.05%
6	22.50±13.69%	33.33±15.51%
7	37.92±17.82%	37.50±16.41%
8	44.17±17.60%	42.08±17.52%
9	14.58±10.93%	22.50±12.46%
10	58.75±19.46%	61.25±17.78%
11	55.00±18.16%	54.58±15.57%
AVG	41.55±15.07%	41.66±14.81%

C.8 Results: OsM method

Throughout the next sections we will report the accuracy results obtained by the OsM method described in section 5.2.1, which corresponds to one of the main contribution of the present work. As mentioned, this approach generates P binary problems, where P corresponds to the number of body parts included in the paradigm (2 problems for the 4-*class* database, and 3 problems for the 8-*class* one intended to determine whether each one of the considered limbs are engaged in the motor imagery or not).

In order to explore different solutions, we have applied this partition technique on both databases in combination with all the feature extraction algorithms that have been presented in

chapter 4.

C.8.1   Results: OsM method by using the CSP algorithm

The first arrangement to be presented for the implementation of the OsM approach is in combination with the CSP algorithm. As mentioned 2 pairs of spatial filters were selected (4 in total). The extracted features from each one of the P binary tasks are subsequently concatenated, which generates input vectors of 12 elements each.

Tables C.23 and C.24 present the accuracy results that were obtained, respectively, for the 4-*class* and the 8-*class* databases after classifying the computed features.

Table C.23: Accuracy results for the OsM method in combination with the CSP algorithm on the 4-class database

	Classification method				
			SVM		
Subject	**LDA**	**QDA**	**Linear**	**Polynomial**	**RBF**
1	47.81±3.73%	41.56±5.14%	50.00±5.86%	47.19±6.64%	40.94±6.80%
2	41.25±5.30%	36.56±5.81%	38.75±6.54%	36.87±5.91%	43.44±2.37%
3	63.44±26.37%	60.31±23.66%	54.06±17.89%	53.44±20.80%	60.00±24.13%
4	61.56±2.58%	56.25±3.06%	55.31±6.95%	48.13±4.62%	59.69±5.63%
5	53.75±4.79%	49.06±6.07%	52.50±3.68%	47.81±4.00%	49.69±5.72%
AVG	53.56±8.55%	48.75±8.75%	50.12±8.18%	46.69±8.39%	50.75±8.93%

Table C.24: Accuracy results for the OsM method in combination with the CSP algorithm on the 8-class database

	Classification method	
Subject	**LDA**	**QDA**
1	36.25±12.44%	30.42±16.31%
2	47.92±18.00%	45.83±19.51%
3	86.25±10.04%	75.00±14.68%
4	45.83±20.32%	37.92±16.57%
5	24.58±14.86%	18.75±12.17%
6	40.83±19.68%	35.00±17.80%
7	40.83±17.96%	30.42±15.98%
8	42.08±15.57%	40.83±14.28%

9	39.17±16.65%	29.58±17.82%
10	61.67±18.26%	54.17±18.95%
11	59.58±14.56%	54.17±17.16%
AVG	47.73±16.21%	41.1±16.48%

C.8.2   Results: OsM method by using the ACSP algorithm

Here we present the results generated by the OsM method in combination with the ACSP algorithm. The process is similar to the one discussed in the previous section, the main difference is that in this case the analytical form of the CSP algorithm explained in section 4.2 is considered. Given that the computed features are defined in the complex space, only results by using linear discriminant analysis classification are presented.

Tables C.25 and C.26 show the accuracy results after classifying the obtained features over the 4-*class* and the 8-*class* databases respectively.

Table C.25: Accuracy results for the OsM method in combination with the ACSP algorithm on the 4-class database

	Classification method	
Subject	**LDA**	**QDA**
1	50.94±10.48%	52.81±10.02%
2	42.81±4.25%	39.06±5.04%
3	62.81±22.94%	55.35±21.27%
4	60.94±2.13%	59.06±7.10%
5	52.50±6.85%	49.06±7.86%
AVG	54.00±9.33%	51.06±10.26%

Table C.26: Accuracy results for the OsM method in combination with the ACSP algorithm on the 8-class database

	Classification method	
Subject	**LDA**	**QDA**
1	35.00±14.83%	30.42±14.56%
2	48.33±14.95%	45.83±18.38%
3	86.67±9.25%	80.42±10.73%
4	47.92±19.44%	37.92±16.24%
5	25.00±17.68%	17.50±12.97%

6	47.08±19.61%	42.50±19.31%
7	42.92±13.41%	35.00±17.18%
8	47.92±18.30%	40.83±17.66%
9	35.83±16.33%	31.67±14.58%
10	62.92±18.71%	60.83±17.90%
11	62.92±18.71%	54.58±14.11%
AVG	49.32±16.47%	43.41±15.78%

C.8.3 Results: OsM method by using the FBCSP algorithm

The following results present the yielded accuracy by the OsM method in combination with the FBCSP algorithm. As mentioned, 2 pairs of CSP filters (4 filters in total) were selected for each one of the 9 frequency bands included in the filter bank. Subsequently, only 4 components are considered; which results in a total number of filters that varies from 4 to 8 depending on whether the selected elements belong to same pair of not (see section 4.4). Finally, the extracted features from each one of the binary classification tasks (2 problems for the 4-*class* database, and 3 for the 8-*class* database) are concatenated into a single vector.

Tables C.27 and C.28 show the performance obtained after classifying the resulting features over the 4-*class* and 8-*class* databases, respectively.

Table C.27: Accuracy results for the OsM method in combination with the FBCSP algorithm on the 4-class database

			Classification method		
				SVM	
Subject	**LDA**	**QDA**	**Linear**	**Polynomial**	**RBF**
1	52.19±6.16%	43.44±8.50%	48.75±6.54%	37.19±5.24%	48.44±6.24%
2	35.63±7.47%	39.69±5.14%	32.50±6.69%	29.69±6.95%	34.69±4.13%
3	68.13±14.77%	67.50±15.65%	54.06±7.53%	33.13±61.25%	61.25±13.73%
4	56.25±4.33%	49.06±5.90%	50.00±6.85%	33.12±9.71%	50.31±4.00%
5	51.56±4.13%	45.00±3.23%	47.81±2.77%	37.81±5.14%	47.50±6.69%
AVG	52.75±7.37%	48.94±7.68%	46.62±6.07%	34.19±6.54%	48.44±6.96%

Table C.28: Accuracy results for the OsM method in combination with the FBCSP algorithm on the 8-class database

Classification method	

Subject	LDA	QDA
1	20.83±11.53%	24.17±14.66%
2	39.17±14.58%	41.25±15.79%
3	82.08±13.80%	89.17±8.42%
4	34.17±16.72%	45.83±17.47%
5	24.58±14.49%	22.08±14.93%
6	41.25±18.90%	34.17±17.96%
7	29.17±17.47%	24.58±13.33%
8	40.83±18.55%	25.00±13.93%
9	31.25±16.00%	27.08±15.07%
10	51.25±17.48%	32.50±14.53%
11	56.25±10.76%	45.42±20.10%
AVG	40.98±15.48%	37.39±15.11%

C.8.4   Results: OsM method by using the FBACSP algorithm

The results shown below were obtained by applying the same method described in the previous section, except that the CSP algorithm was considered in its analytical form. Given that the generated features are defined in the complex space, results in combination with SVM classification for the 4-*class* database are not reported.

Tables C.29 and C.30 show the accuracies obtained for each subject by using linear discriminant analysis classification over the 4-*class* and the 8-*class* databases respectively.

Table C.29: Accuracy results for the OsM method in combination with the FBACSP algorithm on the 4-class database

	Classification method	
Subject	LDA	QDA
1	48.12±8.32%	46.25±6.21%
2	35.00±9.30%	32.19±8.00%
3	69.37±14.31%	69.12±14.01%
4	53.12±6.57%	52.81±3.29%
5	52.19±5.34%	47.19±8.06%
AVG	51.56±8.77%	49.31±7.91%

Table C.30: Accuracy results for the OsM method in combination with the FBACSP algorithm on the 8-class database

	Classification method	
Subject	**LDA**	**QDA**
1	22.50±12.46%	20.00±9.63%
2	35.83±13.82%	27.50±14.08%
3	75.83±15.37%	31.97±9.13%
4	36.25±16.53%	20.42±13.73%
5	21.25±12.78%	20.00±15.95%
6	42.08±14.11%	26.25±15.17%
7	33.75±14.74%	25.83±16.72%
8	35.42±16.44%	24.17±15.02%
9	35.83±13.82%	24.58±16.24%
10	46.67±16.72%	20.00±11.18%
11	52.92±16.64%	21.67±15.37%
AVG	39.85±14.86%	23.85±13.84%

C.8.5   Results: OsM method by using the MDRM approach

In this section we report the performance of the OsM approach in combination with the Minimum Distance to Riemannian Mean algorithm. The implementation of this solution admits different variants. For instance, it is possible to compute the Riemannian distances separating the input elements from each one of the Riemannian class means formulated by the problem (i.e., 4 class means for the 4-*class* database, and 6 class means for the 8-*class* database, given that for each one of the binary subproblems the Riemmanian means are computed over both the C_{ERD} and the C_{IDLE} classes, see section 5.2.1); and to obtain the predicted label from the combination of the shortest distances separating the input data from the means associated to each one of the independent problems. In fact, given the similitude with the MDRM algorithm, this is the approach that we have decided to implement. But another option, which remains to be evaluated, would consist in classifying the computed Riemannian distances with a discriminant linear analysis function, for example, and to concatenate the resulting Euclidean distances into a single P-dimensional vector, where P corresponds to the number of body parts considered by the paradigm, from where the final decision can be inferred by using another classification algorithm.

Results generated by the first approach are presented for the 4-*class* database and the 8-*class* database, respectively, in tables C.31 and C.32.

Table C.31: Accuracy results for the OsM method in combination with the MDRM algorithm on the 4-class database

Subject	**Accuracy**

1	44.69±12.39%
2	31.25±2.28%
3	61.25±5.68%
4	37.50±5.10%
5	49.06±6.72%
AVG	44.75±6.44%

Table C.32: Accuracy results for the OsM method in combination with the MDRM algorithm on the 8-class database

Subject	Accuracy
1	27.08±10.42%
2	39.58±10.93%
3	52.92±13.41%
4	28.33±11.81%
5	20.83±8.89%
6	26.25±11.99%
7	37.08±17.82%
8	32.50±14.53%
9	16.25±10.96%
10	44.58±17.27%
11	50.00±14.31%
AVG	34.13±12.94%

C.8.6   Results: OsM method by using CSP + MDRM approach

The last configuration applied under the OsM approach is the combination of the CSP algorithm with the MDRM method discussed in section 4.6.2.2. Stated briefly, it consists in the same procedure described in the preceding section with the only difference that the covariance matrices, which are used as the EEG descriptors, are computed over the trials projection obtained after applying the CSP algorithm, instead of over the band-pass filtered signal segments (as it is done by the standard MDRM). Note that the CSP filters are computed exactly as it was described in section C.8.1.

Tables C.33 and C.34 show the accuracy results obtained from the 4-*class* and the 8-*class* databases respectively.

Table C.33: Accuracy results for the OsM method in combination with the CSP + MDRM algorithm on the 4-class database

Subject	Accuracy
1	47.19±8.38%
2	34.06±4.00%
3	60.31±21.78%
4	50.62±2.60%
5	47.81±5.72%
AVG	48.00±8.50%

Table C.34: Accuracy results for the OsM method in combination with the CSP + MDRM algorithm on the 8-class database

Subject	Accuracy
1	32.92±15.57%
2	34.17±9.81%
3	67.92±13.00%
4	35.42±15.43%
5	17.50±11.18%
6	31.25±12.61%
7	41.67±13.67%
8	35.83±12.60%
9	17.08±8.98%
10	52.50±14.46%
11	46.25±15.10%
AVG	37.50±12.95%

C.9   Results: HM method

Throughout the next sections results obtained by the HM approach in combination with all the discussed feature extraction and classification algorithms will be presented. Note that hierarchical methods involve only binary problems that are consecutively applied (see sections 5.1.3 and 5.2.2), therefore, classification can be directly applied without using any additional grouping approach. Also note that the HM method can start either by considering first the activity associated with the use of the left hand, or by considering the activity associated with the use of the right hand.

In order to evaluate both variants the two approaches have been implemented for all subjects.

C.9.1   Results: HM method by using the CSP algorithm

To start with, we report the performance of the HM approach in combination with the standard CSP method. Again, we have considered 2 pairs of spatial filters (4 in total) for each one of the binary problems formulated at each stage of the method (2 stages comprising 3 binary problems for the 4-*class* database, and 3 stages containing 7 binary problems for the 8-*class* database).

Tables C.35 and C.36 present, respectively, the accuracy results obtained over the 4-*class* database by analysing first the activity generated by the use of the left hand, and by starting the analysis on the activity generated by the use of the right hand. For the 8-*class* database results by starting to analyze the activity generated by the use of the left hand are shown in table C.37, and in table C.38 when the analysis starts on the activity associated to the use of the right hand.

Table C.35: Accuracy results for the HM method in combination with the CSP algorithm on the 4-class database starting with the left hand

			Classification method		
				SVM	
Subject	LDA	QDA	Linear	Polynomial	RBF
1	49.06±4.83%	48.13±5.64%	50.94±7.93%	45.00±4.56%	47.81±7.59%
2	38.44±3.44%	41.88±3.31%	38.44±3.87%	40.94±4.38%	37.81±3.59%
3	61.25±23.91%	57.19±19.13%	57.81±22.99%	57.81±22.88%	58.44±23.79%
4	55.31±3.29%	52.81±2.13%	54.06±5.44%	52.81±6.80%	54.69±1.57%
5	52.50±6.04%	49.06±8.38%	51.25±5.44%	50.94±9.21%	52.81±8.86%
AVG	51.31±8.30%	49.81±7.72%	50.50±9.55%	49.50±9.56%	50.31±9.08%

Table C.36: Accuracy results for the HM method in combination with the CSP algorithm on the 4-class database starting with the right hand

			Classification method		
				SVM	
Subject	LDA	QDA	Linear	Polynomial	RBF
1	42.81±3.29%	45.00±5.95%	41.56±4.49%	44.37±5.54%	43.13±8.07%
2	39.06±6.95%	38.13±5.82%	38.44±2.95%	41.25±5.68%	40.31±11.06%
3	59.38±22.21%	60.63±22.42%	59.06±20.24%	57.50±20.21%	62.50±21.77%
4	55.31±1.87%	53.44±4.38%	55.00±2.04%	54.06±3.59%	53.44±4.13%
5	55.00±6.29%	53.44±3.87%	54.69±2.58%	50.31±3.87%	53.12±2.98%

AVG	50.31±8.12%	50.12±8.49%	49.75±6.65%	49.50±7.78%	50.50±9.60%

Table C.37: Accuracy results for the HM method in combination with the CSP algorithm on the 8-class database starting with the left hand

	Classification method	
Subject	**LDA**	**QDA**
1	30.83±17.29%	30.83±14.21%
2	48.33±13.02%	46.67±16.72%
3	75.42±14.86%	76.67±14.21%
4	45.42±17.82%	39.58±13.96%
5	27.50±14.08%	27.50±16.54%
6	30.42±17.27%	31.25±14.21%
7	39.17±18.49%	38.75±17.78%
8	41.25±19.74%	35.42±16.77%
9	24.17±16.72%	20.42±15.22%
10	57.92±15.57%	61.67±15.02%
11	52.50±13.69%	52.08±14.71%
AVG	42.99±13.23%	41.89±15.4%

Table C.38: Accuracy results for the HM method in combination with the CSP algorithm on the 8-class database starting with the right hand

	Classification method	
Subject	**LDA**	**QDA**
1	35.00±15.88%	31.67±14.21%
2	46.67±15.02%	44.58±15.29%
3	77.50±13.69%	79.17±13.67%
4	44.58±19.05%	39.17±16.33%
5	25.83±16.72%	25.83±14.28%
6	30.83±16.65%	28.33±13.90%
7	35.83±18.49%	34.58±18.77%
8	42.92±19.88%	38.33±17.04%
9	29.17±13.67%	27.92±13.41%
10	57.08±16.64%	59.17±13.51%
11	53.33±17.35%	52.50±14.46%

AVG	43.52±16.64%	41.93±14.99%

C.9.2   Results: HM method by using the ACSP algorithm

As in the previous sections, we have considered 2 pairs of spatial filters (4 in total) for the implementation of the HM approach in combination with the ACSP algorithm.

Accuracy results are shown in tables C.39 and C.40 for the 4-*class* database. The first table presents the performance of the method when the analysis starts on the activity associated to the use of the left hand, and the second one reports the results obtained by starting to analyze the activity linked to the right hand. Similarly, for the 8-*class* database, results are shown on tables C.41 and C.42 for the cases where the analysis starts on the left hand and right hand respectively. Note that features extracted by the analytical version of the CSP method are defined in the complex space, therefore, classification with the SVM algorithm is not included.

Table C.39: Accuracy results for the HM method in combination with the ACSP algorithm on the 4-class database starting with the left hand

	Classification method	
Subject	**LDA**	**QDA**
1	54.06±4.72%	51.88±7.94%
2	37.81±4.61%	40.94±6.32%
3	61.56±24.71%	61.25±20.41%
4	55.94±1.57%	54.69±5.24%
5	55.94±9.76%	54.37±7.74%
AVG	53.06±9.07%	52.62±9.53%

Table C.40: Accuracy results for the HM method in combination with the ACSP algorithm on the 4-class database starting with the right hand

	Classification method	
Subject	**LDA**	**QDA**
1	53.12±9.16%	45.94±11.34%
2	41.88±3.75%	39.38±1.25%
3	59.69±23.21%	60.00±23.45%
4	57.50±7.57%	58.75±3.39%
5	52.81±7.59%	54.69±4.72%
AVG	53.00±10.26%	51.75±8.83%

Table C.41: Accuracy results for the HM method in combination with the ACSP algorithm on the 8-class database starting with the left hand

Subject	Classification method	
	LDA	**QDA**
1	28.75±15.45%	29.58±12.92%
2	43.75±13.43%	43.33±14.95%
3	77.50±14.83%	55.42±14.19%
4	45.00±22.41%	45.42±17.82%
5	24.58±12.49%	24.58±12.49%
6	35.42±14.34%	35.00±15.19%
7	39.17±18.20%	36.25±18.08%
8	43.33±17.29%	39.58±18.00%
9	30.00±16.93%	31.67±17.90%
10	55.83±17.60%	55.83±14.58%
11	52.08±17.70%	52.50±14.46%
AVG	43.22±16.42%	40.83±15.51%

Table C.42: Accuracy results for the HM method in combination with the ACSP algorithm on the 8-class database starting with the right hand

Subject	Classification method	
	LDA	**QDA**
1	34.17±15.37%	31.67±16.65%
2	44.17±16.33%	43.75±15.66%
3	79.58±13.33%	56.67±17.90%
4	50.00±18.86%	43.33±14.21%
5	27.08±14.71%	26.67±14.95%
6	34.17±15.37%	35.00±13.69%
7	37.50±14.68%	36.67±16.72%
8	45.00±19.31%	36.25±17.17%
9	31.25±14.95%	31.67±15.65%
10	56.25±16.98%	57.92±16.57%
11	55.42±18.48%	52.92±15.98%

	AVG	44.96±16.22%	41.14±16.01%

C.9.3 Results: HM method by using the FBCSP algorithm

In this section we report the performance of the HM method in combination with the FBCSP algorithm. For the implementation of all classification stages, 2 pairs of spatial filters (4 in total) were selected from each one of the models generated over the 9 frequency bands comprised in the filter bank. Subsequently, from all features, only 4 elements were extracted by using the MIBIF algorithm (see sections 4.4 and 4.4.1); which leads to input vectors whose dimension varies between 4 (if the selected elements belong to the same pairs) to 8 (if the selected elements are taken from different pairs).

For the *4-class* database results are shown on tables C.43 and C.44 which present; respectively, the accuracy results by starting to analyze the activity associated to the use of the left hand, and by starting the analysis on the activity linked to the use of the right hand. Similarly, for the *8-class* database, results are shown in table C.45 for the case when the analysis starts on the activity related to the use of the left hand, and in table C.46 when the analysis starts on the activity associated to the use of the right hand.

Table C.43: Accuracy results for the HM method in combination with the FBCSP algorithm on the 4-class database starting with the left hand

			Classification method		
				SVM	
Subject	LDA	QDA	Linear	Polynomial	RBF
1	50.00±6.85%	50.62±4.73%	36.88±2.98%	37.81±4.72%	37.81±3.12%
2	37.50±2.70%	40.94±0.62%	31.87±2.17%	32.81±5.63%	32.50±4.68%
3	65.00±12.99%	57.81±22.88%	37.19±5.90%	36.56±7.86%	36.56±8.32%
4	52.81±4.49%	50.94±5.90%	36.88±3.15%	38.44±2.13%	36.25±3.29%
5	51.88±3.75%	48.44±4.83%	35.94±1.57%	36.88±2.17%	36.25±1.77%
AVG	51.44±6.16%	49.75±7.79%	35.75±3.15%	36.50±4.50%	36.06±4.23%

Table C.44: Accuracy results for the HM method in combination with the FBCSP algorithm on the 4-class database starting with the right hand

			Classification method		
				SVM	
Subject	LDA	QDA	Linear	Polynomial	RBF

1	52.81±8.19%	50.94±3.29%	38.75±4.21%	38.12±4.39%	38.75±4.21%
2	38.12±2.60%	39.06±6.24%	30.00±3.68%	29.69±3.87%	31.87±4.73%
3	65.62±21.71%	62.50±23.32%	37.81±8.56%	37.50±6.69%	37.81±8.62%
4	51.25±7.22%	49.06±4.00%	38.81±1.57%	38.12±2.60%	37.50±1.77%
5	53.44±3.29%	54.69±4.49%	37.50±2.70%	37.50±2.60%	38.44±2.13%
AVG	55.25±8.60%	51.25±8.27%	36.38±4.14%	36.19±3.86%	36.88±4.29%

Table C.45: Accuracy results for the HM method in combination with the FBCSP algorithm on the 8-class database starting with the left hand

	Classification method	
Subject	**LDA**	**QDA**
1	26.25±13.67%	26.67±13.82%
2	37.08±19.83%	38.75±18.38%
3	88.33±9.81%	87.92±12.92%
4	38.33±17.04%	35.00±16.22%
5	21.67±14.28%	20.00±16.28%
6	42.92±18.18%	41.67±16.19%
7	30.00±18.45%	31.67±13.02%
8	42.92±18.77%	42.92±20.68%
9	37.50±16.08%	40.42±16.31%
10	46.25±17.72%	48.33±13.21%
11	56.67±13.82%	57.08±15.98%
AVG	42.49±16.15%	42.77±15.73%

Table C.46: Accuracy results for the HM method in combination with the FBCSP algorithm on the 8-class database starting with the right hand

	Classification method	
Subject	**LDA**	**QDA**
1	26.67±11.71%	28.75±14.74%
2	34.17±15.37%	33.33±19.51%
3	85.83±13.82%	55.00±13.77%
4	38.75±18.08%	40.00±15.88%
5	18.33±12.17%	17.92±11.22%

6	38.75±17.17%	41.25±18.61%
7	31.67±15.65%	31.67±12.60%
8	34.58±16.31%	35.42±16.44%
9	31.67±20.16%	32.08±19.05%
10	47.92±18.88%	50.00±13.53%
11	55.42±15.63%	50.42±13.73%
AVG	40.34±15.90%	37.8±15.37%

C.9.4   Results: HM method by using the FBACSP algorithm

In this section we present the accuracy results obtained by the HM approach in combination with the FBACSP algorithm. Its implementation was similar to the one described in the previous section with the only difference that the CSP algorithm was considered in its analytical form, therefore, given that the resulting features are complex, classification was carried out only by linear discriminant analysis.

Tables C.47 and C.48 present, respectively, the accuracy results obtained over the 4-*class* database by analysing first the activity generated by the use of the left hand, and by starting the analysis on the activity generated by the use of the right hand. For the 8-*class* database results by starting to analyze the activity generated by the use of the left hand are shown in table C.49, and in table C.50 when the analysis starts on the activity associated to the use of the right hand.

Table C.47: Accuracy results for the HM method in combination with the FBACSP algorithm on the 4-class database starting with the left hand

	Classification method	
Subject	**LDA**	**QDA**
1	49.69±8.25%	49.06±6.49%
2	30.00±4.33%	29.38±2.60%
3	60.31±23.70%	60.31±17.72%
4	46.25±3.68%	48.75±1.77%
5	49.06±4.37%	50.00±5.20%
AVG	47.06±8.87%	47.50±6.76%

Table C.48: Accuracy results for the HM method in combination with the FBACSP algorithm on the 4-class database starting with the right hand

	Classification method	

Subject	LDA	QDA
1	45.94±6.40%	48.13±5.99%
2	37.50±6.04%	37.81±7.02%
3	56.88±19.86%	61.88±19.43%
4	50.31±1.88%	47.81±1.87%
5	55.00±9.52%	52.81±8.32%
AVG	49.12±8.74%	49.69±8.53%

Table C.49: Accuracy results for the HM method in combination with the FBACSP algorithm on the 8-class database starting with the left hand

	Classification method	
Subject	**LDA**	**QDA**
1	28.75±14.74%	30.00±17.25%
2	37.50±16.74%	35.42±15.77%
3	87.92±12.49%	53.75±13.59%
4	37.08±17.21%	37.92±16.57%
5	26.25±12.86%	25.42±13.33%
6	39.58±22.76%	42.92±20.15%
7	35.00±15.54%	37.92±15.21%
8	39.58±13.57%	38.75±16.20%
9	35.42±17.08%	31.67±17.29%
10	48.33±18.78%	47.92±16.11%
11	49.58±13.73%	51.67±15.30%
AVG	42.27±15.95%	39.40±16.07%

Table C.50: Accuracy results for the HM method in combination with the FBACSP algorithm on the 8-class database starting with the right hand

	Classification method	
Subject	**LDA**	**QDA**
1	26.67±11.71%	28.75±14.74%
2	34.17±15.37%	33.33±19.51%
3	85.83±13.82%	55.00±13.77%
4	38.75±18.08%	40.00±15.88%

5	18.33±12.17%	17.92±11.22%
6	38.75±17.17%	41.25±18.61%
7	31.67±15.65%	31.67±12.60%
8	34.58±16.31%	35.42±16.44%
9	31.67±20.16%	32.08±19.05%
10	51.67±13.43%	50.00±13.53%
11	50.83±15.02%	50.42±13.73%
AVG	40.27±15.35%	37.80±15.37%

C.9.5   Results: HM method by using the MDRM algorithm

Tables C.51 and C.52 show, respectively, the performance of the HM approach applied in combination with the MDRM algorithm over the 4-*class* and the 8-*class* databases. For each one of the stages formulated by the HM approach two Riemannian means are computed, namely, the one associated to the C_{ERD} class, and the one related to C_{IDLE} class. Subsequently, decisions are made by selecting the groups generating the shortest Riemannian distances between the corresponding class means and the input elements.

Tables C.51 and C.52 show, respectively, the accuracy results over the 4-*class* database when the process starts by analysing the activity generated by the use of the left hand, and by starting the analysis on the activity associated to the use of the right hand. Similarly, for the 8-*class* database, table C.53 presents the results obtained when the analysis starts on the activity generated by use of the left hand, and table C.54, when it starts with the right hand.

Table C.51: Accuracy results for the HM method in combination with the MDRM algorithm on the 4-class database starting with the left hand

Subject	Accuracy
1	40.31±10.92%
2	34.69±3.73%
3	56.87±8.07%
4	47.81±2.37%
5	53.75±4.79%
AVG	46.69±5.97%

Table C.52: Accuracy results for the HM method in combination with the MDRM algorithm on the 4-class database starting with the right hand

Subject	Accuracy

1	44.37±10.13%
2	34.06±2.77%
3	54.69±7.80%
4	47.19±2.13%
5	52.50±6.21%
AVG	46.56±5.81%

Table C.53: Accuracy results for the HM method in combination with the MDRM algorithm on the 8-class database starting with the left hand

Subject	Accuracy
1	28.75±13.99%
2	36.67±12.25%
3	62.08±16.24%
4	35.42±17.08%
5	25.42±12.05%
6	24.17±13.10%
7	39.58±15.43%
8	38.75±17.78%
9	19.58±10.73%
10	50.42±13.73%
11	53.75±15.45%
AVG	37.69±14.35%

Table C.54: Accuracy results for the HM method in combination with the MDRM algorithm on the 8-class database starting with the right hand

Subject	Accuracy
1	26.67±16.97%
2	35.42±12.32%
3	65.83±13.10%
4	36.25±18.67%
5	19.17±12.60%
6	21.25±11.90%
7	32.92±15.57%
8	35.83±16.65%

9	19.58±9.67%
10	47.92±17.40%
11	45.83±15.51%
AVG	35.15±14.58%

C.9.6 Results: HM method by using the CSP + MDRM approach

We have also applied the HM approach in combination with the CSP algorithm preceding the MDRM method, so that the EEG covariance matrices become more discriminative between conditions after the spatial filtering (see section 4.6.2.2).

Tables C.55 and C.56 show, respectively, the performance over the 4-*class* database when the analysis starts on the activity generated by the left hand, and when it starts on the right hand. Similarly, for the 8-*class* database, table C.57 presents the results obtained when the analysis starts on the activity generated by use of the left hand, and table C.58, when it starts with the right hand.

Table C.55: Accuracy results for the HM method in combination with the CSP+MDRM approach on the 4-class database starting with the left hand

Subject	Accuracy
1	50.00±7.36%
2	37.19±4.13%
3	54.69±16.94%
4	54.06±5.98%
5	54.38±6.17%
AVG	50.06±8.12%

Table C.56: Accuracy results for the HM method in combination with the CSP+MDRM approach on the 4-class database starting with the right hand

Subject	Accuracy
1	50.00±8.54%
2	40.31±6.24%
3	61.25±22.48%
4	53.44±2.13%
5	55.63±4.62%
AVG	52.12±8.80%

Table C.57: Accuracy results for the HM method in combination with the CSP+MDRM approach on the 8-class database starting with the left hand

Subject	Accuracy
1	29.58±14.11%
2	39.17±15.99%
3	80.42±13.80%
4	42.50±20.13%
5	25.00±9.28%
6	24.58±12.05%
7	40.83±17.96%
8	43.75±16.33%
9	22.50±15.19%
10	56.25±17.29%
11	52.08±15.77%
AVG	41.51±15.26%

Table C.58: Accuracy results for the HM method in combination with the CSP+MDRM approach on the 8-class database starting with the right hand

Subject	Accuracy
1	32.08±13.80%
2	39.58±10.93%
3	65.00±8.93%
4	47.50±18.10%
5	22.08±14.56%
6	25.42±14.49%
7	37.50±15.74%
8	38.33±13.51%
9	21.67±11.34%
10	54.17±18.38%
11	46.25±11.90%
AVG	39.05±13.79%

C.10 Results: OsHM method

Finally, we report the performance of the OsHM approach, which corresponds to the third contributed method. As mentioned in section 5.2.3, this method is a combination of the OsM and HM approaches; it implements the same binary problems related to the sources associated to the hands, as it was done with the OsM method, and incorporates the four instances associated with the source linked to the feet that were formulated by the HM algorithm in the case of 8-*class* database. Each one of these problems is implemented over the whole set of classes and, subsequently, the resulting features are concatenated to create single vectors, from which classification can be directly achieved. Note the OsHM approach is equivalent to the OsM method when only the 4 classes comprised in the 4-*class* database are considered, therefore, we only report the results obtained over the 8-class database.

C.10.1 Results: OsHM method by using the CSP algorithm

In this section we present the performance of the OsHM method in combination with the CSP algorithm. Given that 2 pairs of spatial filters (4 in total) were considered for each one of the 6 models involved in the procedure, the resulting vectors contain 24 elements each.

Results after applying the algorithm are presented in table C.59.

Table C.59: Accuracy results for the OsHM method in combination with the CSP algorithm on the 8-class database

Subject	Classification method	
	LDA	QDA
1	34.17±13.10%	30.00±16.61%
2	47.08±15.63%	41.25±15.79%
3	82.50±12.97%	77.50±15.54%
4	53.33±17.96%	50.00±21.77%
5	24.58±14.86%	22.50±15.88%
6	42.92±17.27%	42.50±17.25%
7	34.58±14.19%	33.75±13.99%
8	43.50±16.61%	36.67±18.84%
9	39.17±15.99%	38.33±18.55%
10	59.17±17.04%	56.25±17.29%
11	57.50±17.56%	53.33±18.55%
AVG	47.14±15.74%	43.83±17.28%

C.10.2   Results: OsHM method by using the ACSP algorithm

The following results corresponds to the ones generated by the OsHM approach in combination with the ACSP algorithm. Similar to the previous section, we have decided to select 2 pairs of spatial filters (4 in total) for the 6 binary problems formulated by OsHM approach, which generates complex features of 24 elements each. Performance is presented in table C.60.

Table C.60: Accuracy results for the OsHM method in combination with the ACSP algorithm on the 8-class database

Subject	Classification method	
	LDA	**QDA**
1	32.08±15.98%	24.17±12.69%
2	40.42±15.29%	37.50±16.74%
3	81.67±10.75%	74.58±16.24%
4	53.75±23.71%	46.25±20.28%
5	27.08±15.07%	19.17±15.30%
6	44.17±19.35%	43.75±23.61%
7	37.08±16.24%	39.17±16.65%
8	42.92±17.88%	37.92±15.91%
9	37.08±15.57%	36.25±16.20%
10	59.17±13.51%	50.83±16.06%
11	58.75±16.13%	51.67±14.95%
AVG	46.74±16.32%	41.93±16.78%

C.10.3   Results: OsHM method by using the FBCSP algorithm

For each one of the 6 binary problems formulated by the OsHM approach, the FBCSP algorithm was implemented by choosing only 4 elements of the 36 components that result from the concatenation of the 2 pairs of features generated by the models formulated for each one of the 9 frequency bands considered in the filter bank (see sections 4.4 and 4.4.1). For each one of the 6 binary problems, this leads to elements whose dimension varies between 4 (if the selected elements belong to the same pairs) to 8 (if the selected elements are taken from different pairs), which are subsequently concatenated to achieve the corresponding classification.

   Table C.61 presents the accuracy results. Note that results with the linear discriminant analysis when considering different covariance matrices for the involved conditions (i.e., QDA) are missing. This is due to the fact that, in many cases, the features covariance matrices were singular, which means that one or more vectors within the matrix are not linearly independent, and hence, it is not invertible, which is necessary in order to use the linear discriminant analysis

(see section 4.6.1.3).

Table C.61: Accuracy results for the OsHM method in combination with the FBCSP algorithm on the 8-class database

	LDA
Subject	OVO
1	29.17±14.43%
2	33.75±13.19%
3	91.25±8.78%
4	42.08±14.11%
5	19.58±10.73%
6	40.00±18.97%
7	30.83±16.97%
8	47.08±14.56%
9	37.92±17.52%
10	53.33±16.06%
11	60.42±13.16%
AVG	44.13±14.41%

C.10.4 Results: OsHM method by using the FBACSP algorithm

We have also applied the FBACSP algorithm to evaluate the performance of the OsHM approach. Its implementation was similar to the one discussed in the previous section, the only difference is that instead of using the standard CSP algorithm for extracting features within each one of the 9 frequency bands, we have applied its analytical version. Results are shown in table C.62. Note that only results obtained by linear discriminant analysis under the assumption of identical covariance matrices for the involved conditions are presented since, as in the previous section, in many cases their covariance matrices were also singular.

Table C.62: Accuracy results for the OsHM method in combination with the FBACSP algorithm on the 8-class database

	LDA
Subject	OVO
1	29.17±16.52%
2	41.67±14.43%
3	87.08±15.22%
4	47.08±18.18%

5	19.17±14.21%
6	40.00±16.54%
7	35.42±18.30%
8	38.75±13.27%
9	33.33±16.52%
10	51.67±15.65%
11	56.67±16.33%
AVG	42.33±15.92%

C.10.5 Results: OsHM method by using the MDRM algorithm

Similar to what we have discussed in section C.8.5, the implementation of the OsHM approach in combination with the MDRM algorithm admits different variants. However, we have decided to stay as close to the MDRM method, so that labels have been predicted from the combination of the shortest distances separating the input data from the means associated to each one of the 6 binary problems formulated by the OsHM approach. The corresponding accuracy results are shown in table C.63.

Table C.63: Accuracy results for the OsHM method in combination with the MDRM approach on the 8-class database

	Label
Subject	**Accuracy**
1 %	
2	37.92±10.63%
3	60.42±15.77%
4	35.42±15.07%
5	23.33±9.13%
6	24.17±12.69%
7	37.50±18.86%
8	35.42±16.11%
9	15.42±8.49%
10	48.75±14.81%
11	49.17±16.72%
AVG	36.82±13.66%

C.10.6 Results: OsHM method by using the CSP + MDRM approach

We have also applied the OsHM approach in combination with the CSP algorithm preceding the MDRM algorithm as discussed in section 4.6.2.2. Which, in short, consists in the same procedure described in the preceding section with the only difference that the covariance matrices, which are used as the EEG descriptors, are computed over the trials projection obtained after applying the CSP algorithm, instead of over the band-pass filtered signal segments (as it is done by the standard MDRM). Note that the CSP filters are computed exactly as it was described in section C.10.1. Table C.64 presents the correspondig results.

Table C.64: Accuracy results for the OsHM method in combination with the CSP+MDRM on the 8-class database

Subject	Accuracy
1	28.33±12.69%
2	42.50±16.28%
3	74.17±15.37%
4	32.08±17.58%
5	19.17±16.33%
6	28.17±15.37%
7	33.75±15.79%
8	33.33±14.43%
9	15.83±9.25%
10	52.08±16.77%
11	47.08±16.64%
AVG	36.95±15.13%

D

Application Example: Robotic Arm Control

Contents

In order to explore new solutions for overcoming the existing limitations of EEG-based neuro-prostheses, this work has explored new alternatives to control devices requiring a high number of commands. To close such study, in the following section we present an implementation example over a commercial device that affords users with a full 3D control through fourteen predefined motions.

To start with, a description of the robotic arm and the operation scheme, which consists of a smart implementation based on a switching-mode control that has been designed to operate all its functions by using only the eight commands that are obtained trough the paradigm presented in section 3.1, are described.

D.1   Description of the robotic arm

The proposed approach has been motivated by a commercial device model JACO 3-fingers by Kinova™ [1], which affords a full 3D control through fourteen predefined motions; including six reaching movements along the three axis, six rotations around the three axis, and a grasping function. These functions are distributed in three main articulations-like that resemble the forearm, with which the device is able of approaching objects, the wrist, with which it is possible to turn the hand towards any direction, and the fingers, which allows to open and close the hand (see Figure D.1). Regarding the use of the fingers, even when the device allows controlling each pair separately, we have decided to consider them all together. In this way the number of required commands decreases without loss of functionality, since the grasping action by using two or three of the fingers has almost the same effect.
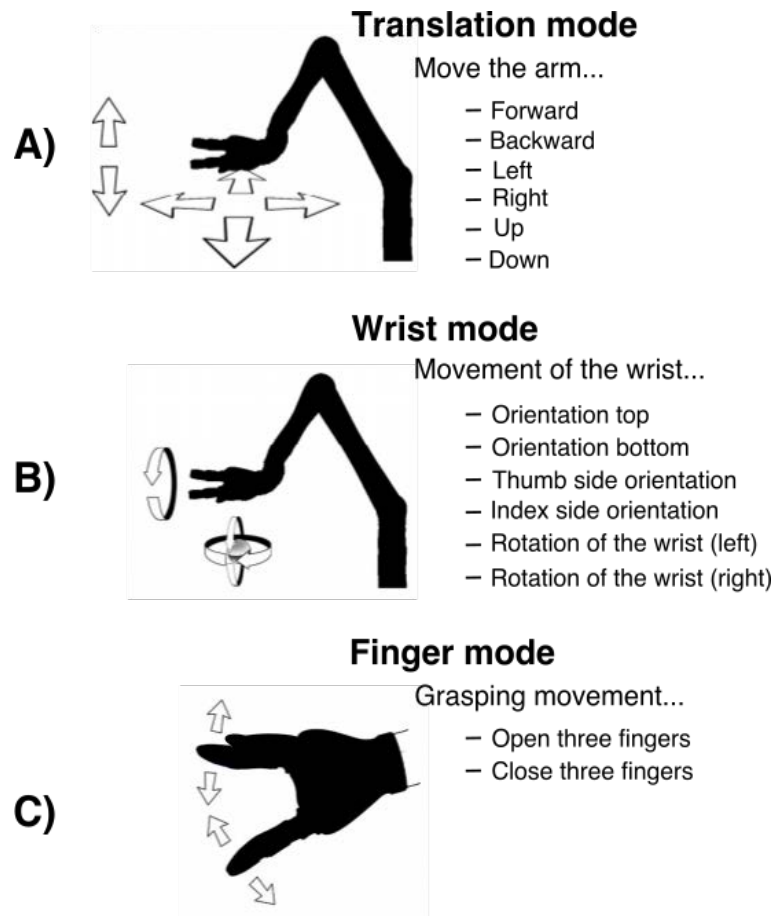
An important aspect of this device, and probably of most of its kind, is that there is a limited range of movements once the robotic arm has reached an extreme position, so that even if the system detects a command demanding a motion that is beyond the possibilities of movement, the corresponding action will not be executed.

D.1.1   Implementation challenges

The appropriate use of this robotic arm implies several challenges that must be addressed from different perspectives. For instance, users' training is one of the most difficult aspects. Even when

---

1. `http://www.kinovarobotics.com/assistive-robotics/products/manipulation/`

Translation mode

Move the arm...

— Forward
— Backward
— Left
— Right
— Up
— Down

A)

Wrist mode

Movement of the wrist...

— Orientation top
— Orientation bottom
— Thumb side orientation
— Index side orientation
— Rotation of the wrist (left)
— Rotation of the wrist (right)

B)

Finger mode

Grasping movement...

— Open three fingers
— Close three fingers

C)

Fig. D.1: 3-axis operation mode. The robotic arm affords a full 3D control through fourteen predefined motions; including six reaching movements along the three axis (A), six rotations around the three axis (B), and a grasping function (C). These functions are distributed in three main articulations-like that resemble the forearm, the wrist, and the fingers. Note that we have decided to use all fingers together, so that the number of required commands decreases without loss of functionality, since the grasping action by using two or three of the fingers has almost the same effect.

powerful feature extraction and classification techniques are implemented, the control of this kind of devices becomes unsatisfactory if users are not able of properly modulating their brain activity. In this sense, the proposed paradigm adds a greater difficulty in comparison to the standard protocols, since the movement imagination of more than one body part at the same time requires a higher degree of coordination. Furthermore, training sessions can be long and tedious for users, specially if the task is not well accomplished, which might cause frustration and make the training process even more difficult. It must be mentioned that this aspect represents one of the main areas of BCI research, which was beyond of the scope of the present project and, therefore, it remains to be addressed as part of the related feature work.

D.1.1.1 Multiple commands

As shown in Figure D.1 this device offers several predefined motions, which on the one hand facilitates the interaction with the subject, but on the other hand demands a considerable number of different brain states to be properly induced and classified. In this regard there are at least two aspects to be considered; first, the number of body parts with which it is possible to offer an intuitive control and whose related sources are separated enough to distinguish the induced activity is limited, and second, the classification task becomes more difficult as the amount of classes increases, which can lead to misclassified responses that hinder the correct use of the device. Because of these considerations we have decided to explore the use of combined motor imageries, since in contrast with the standard scheme, this approach has the advantage of considerably increasing the amount of afforded commands while using the same number of body parts (in order of 2^P compared to P, where P is the number of body parts). However, even when the use of combined movements is considered, there might be not enough commands to control devices as the one we are describing. For instance, in the present work we have decided to consider the use of the left hand, right hand and both feet together. Given that these body parts are intuitively associated with movement. It is worth mentioning that feet are considered together given the proximity of the two sources, which may be difficult to distinguish.

Under this scheme users are afforded only with eight commands if the rest condition is also considered (i.e., left hand, feet, left hand in combination with feet, right hand, both hands together; right hand in combination with feet; both hands together with feet, and rest), which are not sufficient to control the fourteen available motions. To overcome the insufficiency of commands we have designed a smart switching-mode scheme that allows controlling different actions by using the same command.

D.1.1.1.1 Operation scheme

As mentioned, the combination of the three body parts included in the paradigm together with the rest condition, afford users with eight different commands. However, since it is essential to keep the rest condition unassigned to any movement, only seven classes remain to control the device. To this end, an implementation based on three modes has been proposed: one mode for controlling the arm position (see Figure D.1-A), another one for the wrist (see Figure D.1-B), and the third one for opening and closing the fingers (see Figure D.1-C). Each mode assigns a specific class to a specific motion. The combination of the three motor imageries (both hands and feet) has been established to be used as a switch that allows changing from one mode to the next one (arm, wrist, fingers, arm...). In this way it is possible to manage different actions by using the same mental command to control the fourteen available motions (see Figure D.2).

Fig. D.2: Arm control. The different MIs that users were asked to perform are shown in the first column with their associated cue stimuli in the second one. The last three columns present the different robotic movements associated to each mode. Note that the rest condition is not linked to any action, and that motor imagery including the combination of all body parts represents a switch to change from one mode to another.

D.1.1.2 Online setup

One of the interest of using motor imagery for controlling robotic devices is the benefit of providing users with a self-paced paradigm, which makes sense only if the implementation is built upon an online setup (i.e., commands are translated in real time).

In order to implement an online setup there are several aspects to be addressed; perhaps the most important one consists in achieving an accurate classification that delivers results as quickly as possible, so that users are provided with an almost real-time feedback. In this sense, the present work has focused in the implementation of feature extraction methods that favor classification performance from a multilabel basis, so that more discriminative features are generated and computational costs are optimized. There are other aspects to be considered for building online applications, especially if the goal is to use them on a daily basis in out-of-the-lab environments. These aspects are to be addressed as part of the feature work of this project. In the two following sections we mention some characteristics that can favor the implementation of an online setup in combination with the proposed feature extraction methods.

D.2   3D control of a robotic arm, implementation example

The following session presents the implementation of the One-step Multilabel method in combination with the CSP algorithm and LDA classification for controlling the aforementioned robotic

arm. To this end, the user that had shown the most discriminative ERD/ERRS patterns (subject 3), was asked to repeat a recording session (see figure D.3). Such a session comprised three runs of ten trials per class. After the first run, the registered EEG signals were used to train a classification model in order to provide feedback during the second run. Similarly, after the second run, both sets were used to train a new model to provide feedback in the third and last run. In this way, by training the model with data belonging to different runs, it is possible to emulate an online setup, since the variability of data is already significant.
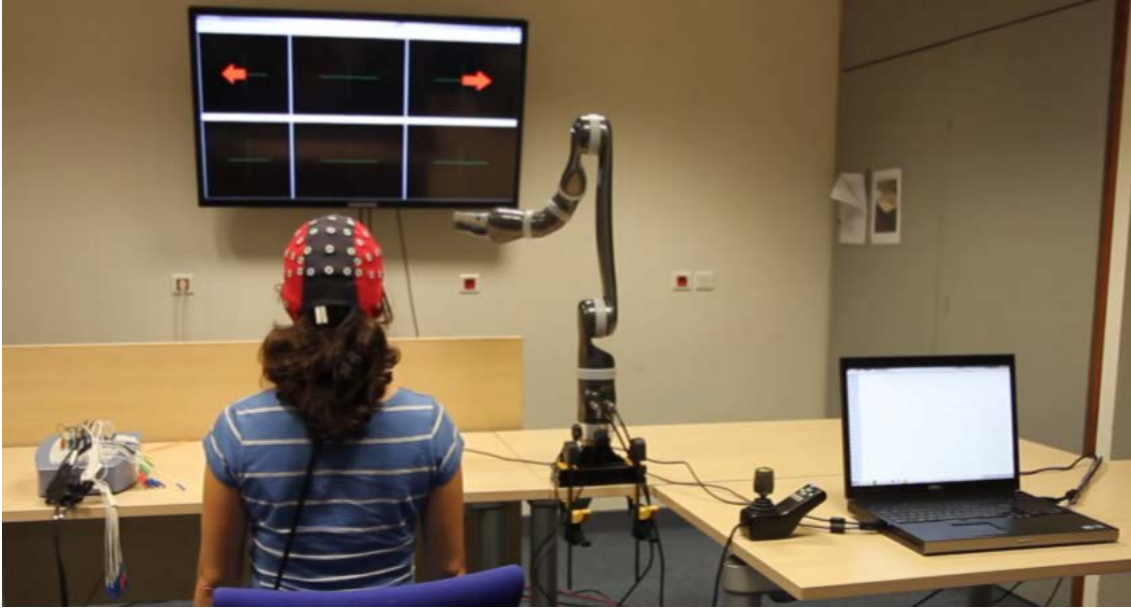
Fig. D.3: Implementation example. The window (displayed on the big screen) is divided in an upper and lower panel. The one in the top shows the stimulation corresponding to the task that the user is asked to perform at each trial. The lower one shows the prediction.

Accuracy after the second run was 78.5% and 84.75% for the third run. Subsequently, an offline analysis was carried out in order to evaluate the system performance by using a 3-fold cross validation for which the test set was, at each fold, all trials belonging to one run. Table D.1 shows the obtained results.

Table D.1: Unpaired T-test to evaluate significance level at 5% of the different classes at electrode C4

Run	Accuracy
1	76.25%
2	81.25%
3	86.25%

In addition, in order to compare the specificity of the OsM method, figure D.4 presents a

graphical representation of the the mean confusion matrices.
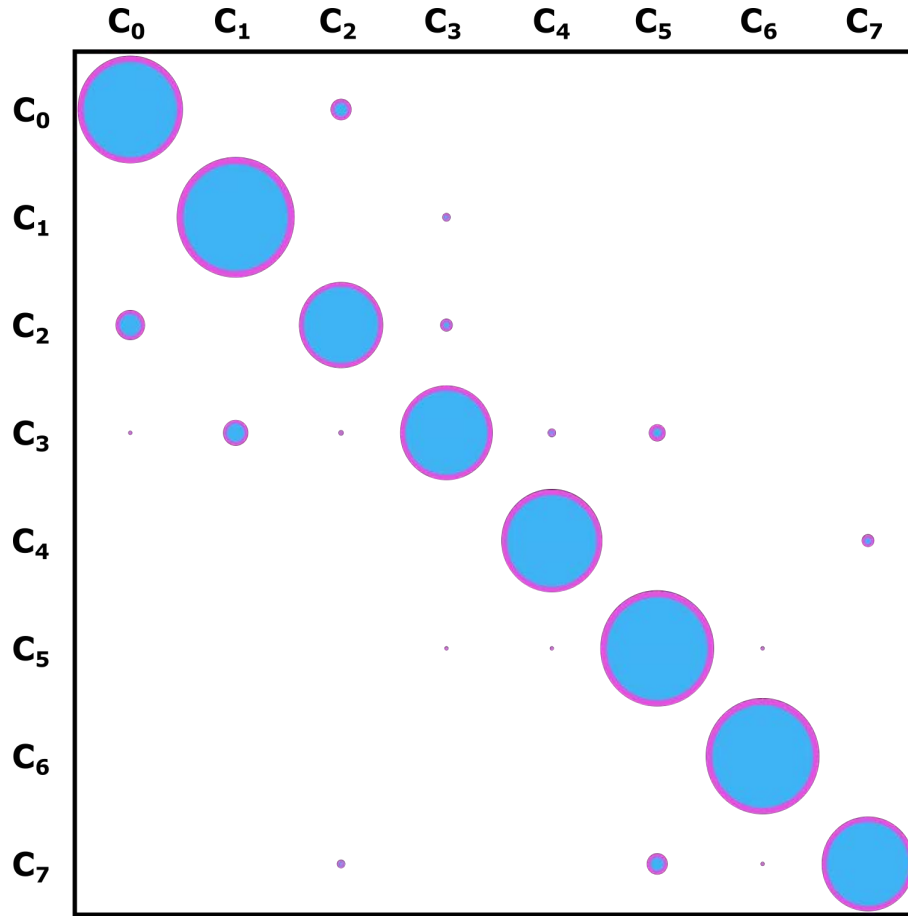
Fig. D.4: Confusion matrices obtained by the OsM approach for subject 3 during an online setup. Note that the surface of each element represents the sum of predictions generating the same label for the trials of each class.

Bibliography

[1] J. R. Wolpaw and E. W. Wolpaw, *BCI: Principles and Practice*. Oxford University Press, 2012.

[2] L. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces – a review.", *Sensors 12, 1211–1279*, 2012.

[3] G. Pfurtscheller, "Event-related synchronization (ERS): an electrophysiological correlate of cortical areas at rest", *Electroencephalogr Clin Neurophysiol*, vol. 83, pp. 62–69, 1992.

[4] S. Fok, R. Schwartz, M. Wronkiewicz, C. Holmes, J. Zhang, T. Somers, D. Bundy, and E. Leuthardt, "An EEG-based brain computer interface for rehabilitation and restoration of hand control following stroke using ipsilateral cortical physiology", *33rd Annual International Conference of the IEEE EMBS*, 2011.

[5] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain–computer communication", *Proc. IEEE*, vol. 89, no. 7, pp. 1123–1134, 2001.

[6] C. Neuper and G. Pfurtscheller, "Evidence for distinct beta resonance frequencies in human EEG related to specific sensorimotor cortical areas", *Clin. Neurophysiol.*, vol. 5, 2001.

[7] G. Pfurtscheller, C. Neuper, C. Brunner, and F. Lopes da Silva, "Beta rebound after different types of motor imagery in man", *Neuroscience Letters*, 2005.

[8] C. Toro, G. Deuschl, R. Thatcher, S. Sato, C. Kufta, and M. Hallett, "Event-related desynchronization and movement-related cortical potentials on the ECoG and EEG", *Electroencephalogr Clin Neurophysiol*, 1994.

[9] C. Babiloni, F. Babiloni, F. Carducci, F. Cincotti, G. Cocozza, C. D. Percio, D. V. Moretti, and P. M. Rossini, "Human cortical electroencephalography (EEG) rhythms during the observation of simple aimless movements: a high-resolution EEG study", *Neuroimage*, vol. 17, pp. 559–72, 2002.

[10] P. Avanzini, M. Fabbri-Destro, R. Dalla Volta, E. Daprati, G. Rizzolatti, and G. Cantalupo, "The dynamics of sensorimotor cortical oscillations during the observation of hand movements: an EEG study", *PLoS One*, vol. 7, no. 5, 2012.

[11] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles", *Clin Neurophysiol*, vol. 110, no. 11, pp. 1842–57, 1999.

[12] G. Pfurtscheller, A. Stancàk, and G. Edlinger, "On the existence of different types of central beta rhythms below 30 hz", *Clin Neurophysiol*, vol. 102, pp. 316–325, 1997.

[13] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multi-class brain computer interface classification by riemannian geometry", *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 920–928, 2012.

[14] G. Dornhege, B. Blankertz, G. Curio, and G. Müller, "Increase information transfer rates in BCI by CSP extension to multi-class", *Advances in neural information processing systems,Proceedings of the 2003 conference; Seventeenth Annual Conference on Neural Information Processing Systems (NIPS)*, vol. 16, 2003.

[15] G. Dornhege, B. Blankertz, G. Curio, and K. Müller, "Boosting bit rates in noninvasive eeg single-trial classifications by feature combination and multiclass paradigms", *IEEE Transactions on Biomedical Engineering*, vol. 6, no. 51, 2004.

[16] Y. Weibo, Q. Shuang, Q. Hongzhi, Z. Lixin, W. Baikun, and M. Dong, "Eeg feature comparison and classification of simple and compound limb motor imagery", *IEEE Transactions on neural systems and rehabilitation engineering*, 2010.

[17] A. S. Royer, A. J. Doud, M. Rose, and B. He, "(EEG) control of a virtual helicopter in 3-dimensional space using intelligent control strategies", *Journal of NeuroEngineering and Rehabilitation*, no. 10, 2013.

[18] Z. Zhou, B. Wan, D. Ming, and H. Qi, "A novel technique for phase synchrony measurement from the complex motor imaginary potential of combined body and limb action", *J Neural Eng*, vol. 7, 11 Pp, 2010.

[19] R. Scherer, S. Zanos, E. Edwards, K. Miller, J. Ojemann, and R. Rao, "Can synergistic finger movements be represented as the superposition of individual finger movements?", *BCI meeting 2010, Asilomar, California, USA.*, 2010.

[20] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanaba, and K. Müller, "Optimizing spatial filters for robust EEG single-trial analysis [revealing tricks of the trade]", *IEEE SIGNAL PROCESSING MAGAZINE*, 2008.

[21] C. Lindig-León and Bougrain, "Comparison of sensorimotor rhythms in EEG signals during simple and combined motor imageries over the contra and ipsilateral hemispheres", in *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Milan Italy (EMBC2015)*, 2015.

[22] C. Lindig-León and L. Bougrain, "A multilabel classification method for detection of combined motor imageries", in *2015 IEEE international conference on systems, man, and cybernetics (SMC2015)*, 2015.

[23] S. Mason and G. Birch, "Temporal control paradigms for direct brain interfaces -rethinking the definition of asynchronous and synchronous.", *Paper presented in HCI International, Las Vegas.*, 2005.

[24] S. Mason, G. Birch, A. Bashashati, M. Fatourechi, K. Navarro, and G. Birch, "A comprehensive survey of brain interface technology designs.", *Ann Biomed Eng*, vol. 35, pp. 137–169, 2007.

[25] J. Mak and J. Wolpaw, "Clinical applications of brain-computer interfaces: current state and future prospects", *IEEE Reviews in Biomedical Engineering*, vol. 2, pp. 187–199, 2009.

[26] F. Lotte, R. Scherer, and A. Lécuyer, "A:8 teaching the bci skill: feedback and human training approaches", *BCI meeting 2013 workshop*, 2013.

[27] J. R. Wolpaw, "Memory in neuroscience: rhetoric versus reality.", *Behav. Cognit Neurosci Rev*, vol. 1, pp. 130–163, 2002.

[28] T. Zander and C. Kothe, "Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general.", *J Neur Engin*, vol. 8, no. 1, 5pp, 2011.

[29] B. Graiman, B. Allison, and G. Pfurtscheller, "Brain-computer interfaces", *Berlin: Springer*, p. 21, 2010.

[30] G. Müller-Putz, R. Scherer, G. Pfurtscheller, and R. Rupp, "Brain-computer interfaces for control of neuroprostheses: from synchronous to asynchronous mode of operation", *Biomed Technik*, vol. 2, no. 51, pp. 57–63, 2006.

[31] T. Lal, T. Hinterberger, G. Widman, M. Schroeder, J. Hill, W. Rosenstiel, C. Elger, B. Schölkopf, and N. Birbaumer, "Methods towards invasive human brain computer interfaces", *Advances in Neural Information Processing System*, no. 17, pp. 737–744, 2005.

[32] D. Nair, R. Burgess, C. McIntyre, and H. Lüders, "Chronic subdural electrodes in the management of epilepsy", *Clin. Neurophysiol.*, vol. 119, 2008.

[33] C. Roy and C. Sherrington, "On the regulation of the blood-supply of the brain", *He Journal of Physiology*, vol. 11, pp. 85–117, 1890.

[34] S. C. Bunce, M. Izzetoglu, K. Izzetoglu, B. Onaral, and K. Pourrezaei, "Functional near-infrared spectroscopy", *IEEE Engineering in Medicine and Biology Magazine*, no. 25, pp. 54–62, 2006.

[35] D. Cohen, "Magnetoencephalography: detection of the brain's electrical activity with a superconducting magnetometer", *Science*, no. 175, pp. 664–666, 1972.

[36] Z. M. Kisley M. A.and Cornwell, "Gamma and beta neural activity evoked during a sensory gating paradigm: effects of auditory, somatosensory and cross-modal stimulation", *Clin. Neurophysiol.*, vol. 117, 2006.

[37] S. Matsuoka, "Theta rhythms: state of consciousness", *Brain Topography*, vol. 3, pp. 203–208, 1990.

[38] G. E. Chatrian, L. Bergamini, and M. Dondey, "A glossary of terms most commonly used by clinical electroencephalographers", *Electroencephalographic Clinical Neurophysiology*, vol. 37, pp. 538–553, 1974.

[39] J. Baumeister, T. Barthel, K. Geiss, and M. Weiss, "Influence of phosphatidylserine on cognitive performance and cortical activity after induced stress", *Nutritional Neuroscience*, vol. 11, no. 3, pp. 103–110, 2008.

[40] S. Baker, "Oscillatory interactions between sensorimotor cortex and the periphery", *Current Opinion in Neurobiology*, vol. 17, no. 6, pp. 649–655, 2007.

[41] E. Lalo, T. Gilbertson, L. Doyle, V. Di Lazzaro, B. Cioni, and P. Brown, "Phasic increases in cortical beta activity are associated with alterations in sensory processing in the human", *Experimental brain research*, vol. 117, no. 1, pp. 137–145, 2007.

[42] E. M. Whitham, K. J. Pope, S. P. Fitzgibbon, T. Lewis, C. R. Clark, S. Loveless, M. Broberg, A. Wallace, D. DeLosAngeles, P. Lillie, A. Hardy, R. Fronsko, A. Pulbrook, and J. O. Willoughby, "Scalp electrical recording during paralysis: quantitative evidence that EEG frequencies above 20 hz are contaminated by EMG", *Clinical Neurophysiology*, vol. 118, pp. 1877–1888, 2007.

[43] E. Niedermeyer and F. L. da Silva, "Electroencephalography: basic principles, clinical applications, and related fields", in. Philadelphia , Lippincott Williams & Wilkins, 2005.

[44] G. Klem, H. Lüders, H. Jasper H., and C. Eldger, "The ten-twenty electrode system of the international federation", *Electroencephalogr Clin Neurophysiol Suppl*, 1999.

[45] "American electroencephalographic society. guideline thirteen: guidelines for standard electrode position nomenclature", *Clin. Neurophysiol.*, 1994.

[46] R. Oostenveld and P. Praamastra, "The 5% electrode system for high-resolution EEG and ERP measurements", *Clin. Neurophysiol.*, 2001.

[47] D. Tucker, "Spatial sample of head electrical fields: the geodesic network net", *Electroencephalogr Clin Neurophysiol*, 1993.

[48] L. E. Miller and N. Hatsopoulos, "Brain-computer interfaces, principles and practice", in. Oxford university press, 2012, ch. Neuronal activity in motor cortex and related areas, pp. 15–43.

[49] G. Fritsch and E. Hitzig, "Über der elektrische erregbarkeit des grosshirns", *Arch Anat Physiol Leipzig*, vol. 37, pp. 330–332, 1870.

[50] W. Penfield and E. Boldrey, "Somatic motor and sensory representation in the cerebra cortex of man as studied by electrical stimulation", *Brain*, vol. 60, pp. 389–443, 1937.

[51] W. Penfield, "Some mechanisms of consciousness discovered during electrical stimulation of the brain", *Proc Natl Acad Sci, USA*, vol. 44, pp. 51–66, 1958.

[52] G. Pfurtscheller and F. Lopes da Silva, "Niedermeyer's electroencephalography", in. Wolters Kubler, 2011, ch. EEG event-related desynchronization ERD and event-related synchronization ERS, pp. 935–948.

[53] G. Pfurtscheller and A. Aranibar, "Event-related cortical desynchronization detected by power measurements of scalp EEG", *Electroencephalogr Clin Neurophysiol*, vol. 42, pp. 817–826, 1977.

[54] W. van Winsum, J. Sergeant, and R. Gueze, "The functional significance of event- related desynchronization of alpha rhythms in attentional and activating tasks", *Electroenceph clin Neurophysiol*, vol. 58, pp. 519–524, 1984.

[55] G. Pfurtscheller and A. Berghold, "Patterns of cortical activation during planning of voluntary movement", *Electroencephalogr Clin Neurophysiol*, vol. 72, pp. 250–258, 1989.

[56] W. Klimesch, G. Pfurtscheller, W. Mohl, and H. Schimke, "Event-related desynchronization, ERD-mapping and hemispheric differences for words and numbers", *J Psychophysiol*, vol. 8, pp. 297–308, 1990.

[57] W. Mohl and G. Pfurtscheller, "The role of the right parietal region in a movement time estimation task", *NeuroReport*, vol. 2, pp. 309–312, 1991.

[58] G. Pfurtscheller, "Synergetics of rhythms in biological systems", in. H. Haken and H. Koepchen (Eds.), Springer, Berlin, 1991, ch. EEG-Rhythms: event-related desynchronization and synchronization, pp. 289–296.

[59] A. Androulidakis, P. Mazzone, V. Litvak, W. Penny, M. Dileone, L. Gaynor, S. Tish, and P. Di Lazzaro V.and Brown, "Oscillatory activity in the penduculopontine area of patients with parkinson's disease", *Exp. Neurol.*, vol. 211, pp. 59–66, 2008.

[60] F. Klostermann, V. Nikulin, A. Kuhn, F. Marzinzink, M. Wahl, A. Pogosyan, A. Kupsch, G. Schneider, P. Brown, and G. Curio, "Task-related differential dynamics of EEG alpha- and beta-band synchronization in cortico-basal motor structures", *Eur J Neurosci*, vol. 25, pp. 1604–1615, 2007.

[61] D. Williams, M. Tijssen, G. van Bruggen, G. Bosh, A. Insola, V. Lazzaro, P. Mazzone, A. Oliviero, A. Quartarone, H. Speelman, and P. Brown, "Dopamine-dependent changes in the functional conectivity between basal ganglia and cerebral cortex in humans", *Brain*, vol. 125, pp. 1558–1569, 2002.

[62] C. Zich, S. Debener, M. De Vos, S. Frerichs, S. Maurer, and C. Kranczioch, "Lateralization patterns of covert but not overt movements change with age: an EEG neurofeedback study", *NeuroImage*, vol. 116, pp. 80–91, 2015.

[63] B. Blankertz, F. Losch, M. Krauledat, G. Dornhege, G. Curio, and K. Müller, "Berlin brain-computer interface: accurate performance from first-session in BCI-naïve subjects", *IEEE Trans. Biomed. Eng*, vol. 55, no. 10, pp. 2452–2462, 2008.

[64] B. Blankertz, C. Sannelli, S. Halder, E. Hammer, A. Kübler, K.-R. Müller, T. Dickhaus, and et al., "Neurophysiological predictor of SMR-based bci performance", *NeuroImage*, vol. 51, no. 4, pp. 1303–1309, 2008.

[65] S. Fazli, J. Mehnert, J. Steinbrink, G. Curio, A. Villringer, K.-R. Müller, and B. Blankertz, "Enhanced performance by a hybrid nirs-eeg brain computer interface", *NeuroImage*, vol. 59, no. 1, pp. 519–529, 2012.

[66] C. Guger, G. Edlinger, W. Harkam, I. Niedermayer, and G. Pfurtscheller, "How many people are able to operate an eeg-based brain–computer interface (BCI)?", *IEEE Trans. Neural. Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 145–147, 2003.

[67] C. Neuper and G. Pfurtscheller, "Event-related desynchronization", in. Handbook of electroencephalography and clinical neurophysiology, Elsevier, 1999, ch. Motor imagery and ERD, pp. 303–325.

[68] C. Zich, S. Debener, C. Kranczioch, M. Bleichner, I. Gutberlet, and M. De Vos, "Real-time EEG feedback during simultaneous EEG-fMRI identifies the cortical signature of motor imagery", *NeuroImage*, vol. 114, pp. 438–447, 2015.

[69] L. Leocani, C. Toro, P. Manganotti, P. Zhuang, and H. M., "Event-related coherence and event-related desynchronization/synchronization in the 10 hz and 20 hz EEG during self-paced movements", *Electroencephalogr Clin Neurophysiol*, vol. 104, pp. 199–206, 1997.

[70] G. Pfurtscheller, "Graphical display and statistical evaluation of event-related desynchronization (ERD)", *Electroencephalogr Clin Neurophysiol*, vol. 43, pp. 757–760, 1977.

[71] G. Pfurtscheller, K. Zalaudek, and C. Neuper, "Event-related beta synchronization after wrist, finger and thumb movement", *Electroencephalogr Clin Neurophysiol*, vol. 102, pp. 154–160, 1998.

[72] M. Taniguchi, A. Kato, N. Fujita, M. Hirata, H. Tanaka, T. Kihara, H. Ninomiya, N. Hirabuki, H. Nakamura, S. Robinson, D. Cheyne, and T. Yoshimine, "Movement-related desynchronization of the cerebral cortex studied with spatially filtered magnetoencephalography", *NeuroImage*, vol. 12, pp. 298–306, 2000.

[73] J. Kaiser, R. Ulrich, and W. Lutzenberger, "Dynamics of sensorimotor cortex activation to spatial sounds precueing ipsi- versus contralateral manual responses", *Cognit Brain Res*, vol. 17, pp. 573–583, 2003.

[74] D. Serrien, "Coordination constraints during bimanual versus unimanual performance conditions", *Neuropsychologia*, vol. 46, pp. 419–425, 2008.

[75] A. Stancàk, J. Svoboda, R. Rachmanova, J. Vrana, J. Kralik, J. Tintera, and G. Pfurtscheller, "Desynchronization of cortical rhythms following cutaneous stimulation: effects of stimulus repetition and intensity, and of the size of corpus callosum", *Clin Neurophysiol*, vol. 114, pp. 1936–1947, 2003.

[76] L. Qin and B. He, "A wavelet-based time–frequency analysis approach for classification of motor imagery for brain–computer interface applications", *J Neural Eng*, vol. 2, pp. 65–72, 2005.

[77] M. Neuper C.and Wortz and G. Pfurtscheller, "ERD/ERS patterns reflecting sensorimotor activation and deactivation", *Prog Brain Res*, vol. 159, pp. 211–222, 2006.

[78] F. Lopes da Silva, "Electroencephalography: basic principles, clinical applications and related fields", in. Baltimore: Williams and Wilkins, 1993, ch. Computer-assisted EEG diagnosis: pattern recognition and brain mapping. Pp. 1063–1086.

[79] G. Pfurtscheller, C. Neuper, and J. Berger, "Source localization using event-related desynchronization (ERD) within the alpha band", *Brain Topogr*, vol. 6, no. 4, pp. 269–275, 1994.

[80] J. Höhne and B. Blankertz, "BBCI toolbox-offline analysis", *BBCIWinter school*, no. 2, pp. 145–147, 2014.

[81] B. Sayers, A. Mc, H. Beayley, and W. Henshall, "The mechanism of auditory evoked eeg responses", *Nature*, vol. 247, pp. 481–483, 1974.

[82] P. Vijn, B. van Dijk, and H. Spekreijse, "Visual stimulation reduced eeg activity in man", *Brain Res*, vol. 550, pp. 49–53, 1991.

[83] L. da Silva F.H., "Neural mechanisms underlying brain waves: from neural membranes to networks", *Electroencephalogr Clin Neurophysiol*, vol. 79, pp. 81–93, 1991.

[84] W. Singer, "Synchronization of cortical activity and its putative role in information processing and learning", *Ann Rev Physiol*, vol. 55, pp. 349–374, 1993.

[85] Y. Renard, F. Lotte, G. Gibert, M. Congedo, and e. a. E. Maby, "Open- vibe: an open-source software platform to design, test and use brain-computer inter- faces in real and virtual environments", *Teleoperators and Virtual Environments, MIT press*, 2010.

[86] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. R. Müller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schlögl, C. Vidaurre, S. Waldert, and B. Blankertz, "Review of the bci competition iv", *Front Neurosci*, vol. 6, p. 55, 2012. DOI: `10.3389/fnins.2012.00055`.

[87] M. C. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[88] M. Taboga, *Lectures on probability and statistics*. 2010.

[89] P. Mahalanobis, "On the generalised distance in statistics", *Proceedings of the National Institute of Sciences of India*, vol. 2, no. 1, pp. 49–55, 1936.

[90] K. Ang, Z. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b", *Frontiers in Neuroscience*, vol. 6, 2012.

[91] O. Falzon, K. Camilleri, and J. Muscat, "The analytic common spatial patterns method for EEG-based BCI data", *Journal of Neural Engineering*, vol. 9, 2012.

[92] L. F. Nicolas-Alonso, R. Corralejo, D. Álvarez, and Hornero, "Analytic common spatial pattern and adaptive classification for multiclass motor imagery-based BCI", in *6th Annual international IEEE-EMBS conference on neural engineering*, 2013.

[93] A. Jain, R. Duin, and M. Jianchang, "Statistical pattern recognition: a review", *IEEE Trans. Pattern Anal.*, no. 22, pp. 4–37, 2000.

[94] B. Blankertz, G. Dornhege, M. Krauledat, K. Müller, and G. Curio, "The non-invasive berlin brain-computer interface: fast acquisition of effective performance in untrained subjects", *Neuroimage*, no. 37, pp. 539–550, 2007.

[95] Z. Ang K.and Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface", *Proceedings of the IEEE International Joint Conference on Neural Networks, Hong Kong*, 2008.

[96] E. Parzen, "On estimation of a probability density function and mode", *Ann. Math. Stat.*, vol. 33, pp. 1065–1076, 1962.

[97] A. Ziehe, P. Laskov, G. Nolte, and K. Müeller, "A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation", *J Machine Learning Research*, vol. 5, pp. 777–800, 2004.

[98] M. Grosse-Wentrup and M. Buss, "Multi-class common spatial patterns and information theoretic feature extraction", *IEEE Transactions on Biomedical Engineering*, vol. 8, no. 55, 2008.

[99] Y. Yang, S. Chevallier, J. Wiart, and I. Bloch, "Automatic selection of the number of spatial filters for motor-imagery BCI", *20th European symposium on artificial neural networks, computational intelligence and machine learning (ESANN 2012)*, pp. 109–114, 2012.

[100] P. McCullagh and J. Nelder, "Generalized linear models", in. Chapman and Hall, 1989.

[101] V. Vapnik, *Estimation of Dependences Based on Empirical Data: Springer Series in Statistics.* Springer-Verlag New York, 1982.

[102] O. Chapelle, "Training a support vector machine in the primal", *Neural Comput*, vol. 19, no. 5, pp. 1155–1178, 2007.

[103] B. Sch olkop and A. Smola, *Learning with Kernels : Support Vector Machines, Regularization, Optimization, and Beyond.* The MIT Press, 1st editio, 2002.

[104] A. Aizerman, E. Braverman, and L. Rozoner, "Theoretical foundations of the potential function method in pattern recognition learning", *Automation and Remote Control*, vol. 25, pp. 821–837, 1964.

[105] A. Barachant, "Covariance toolbox", *Https://github.com/alexandrebarachant*, 2013.

[106] H. Karcher, "Riemannian center of mass and mollifier smoothing", *Communications on Pure and Applied Mathematics*, vol. 30, no. 5, pp. 509–541, 1977.

[107] P. Fletcher and S. Joshi, "Principal geodesic analysis on symmetric spaces: statistics of diffusion tensors", *Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis*, 2004.

[108] V. Vapnik, *Statistical learning theory*. Wiley, 1998.

[109] C. Lindig-León, N. Gayraud, L. Bougrain, and M. Clerc, "Hierarchical classification using riemannian geometry for motor imagery based bci systems", in *BCI meeting 2016, Asilomar, California, USA.*, 2016.

[110] G. Tsoumakas and I. Katakis, "Multilabel classification: an overview", *International Journal of Data Warehousing & Mining*, vol. 3, no. 3, pp. 1–13, 2007.

[111] G. Tsoumakas, E. Spyromitros-Xioufis, J. Vilcek, and I. Vlahavas, "Mulan: a java library for multi-label learning", *Journal of Machine Learning Research*, vol. 12, pp. 2411–2414, 2011.

[112] H. Suk and S. Lee, "Subject and class specific frequency bands selection for multiclass motor imagery classification", *Journal of Imaging Systems and Technology*, vol. 21, pp. 123–130, 2011.

[113] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and other kernel-base learning methods*. Cambridge University Press, 2000.

[114] S. Haufe, F. Meinecke, K. Görgen, S. Dähane, J.-D. Haynes, B. Blankertz, and B. F., "On the interpretation of weight vectors of linear models in multivariate neuroimaging", *NeuroImage*, vol. 87, pp. 96–110, 2014.

[115] C. Vidaurre and B. Blankertz, "Towards a cure for bci illiteracy", *Brain Topography*, vol. 23, pp. 194–198, 2010.

[116] S. Brandl, J. Höhne, and K.-R. Müller, "Bringing bci into everyday life: motor imagery in a pseudo realistic environment", in *7th International IEEE EMBS conference on neural engineering (NER15)*, 2015.

[117] C. Jeunet, E. Jahanpour, and F. Lotte, "Why standard brain-computer interface (BCI) training protocols should be changed: an experimental study", *Journal of Neural Engineering*, vol. 13, no. 3, 2016.

[118] C. Lindig-León, L. Bougrain, and S. Rimbert, "Alpha rebound improves on-line detection of the end of motor imageries", in *7th International IEEE EMBS conference on neural engineering (NER15)*, 2015.

[119] B. Blankertz, C. Schäfer, G. Dornhege, and G. Curio, "Single trial detection of eeg error potentials: a tool for increasing BCI transmission rates", *Artificial Neural Networks -ICANN*, no. 4, pp. 1303–1309, 2002.