
# Evaluation of different training strategies for motor-imagery brain-machine interfaces

Thesis by

David Manetta

Thesis supervisor

Prof. Luca Tonin

Master of Bioengineering for Neuroscience

University of Padua

Academic Year 2021/2022

17/10/2022

# ABSTRACT

Evaluation of different training strategies for motor-imagery brain-machine interfaces

David Manetta

Last years have seen an increasing interest in the field of human-machine interaction (HMI), especially thanks to new methods to acquire signals and new improvements on processing algorithms. For people affected by deabilitant diseases that make movement difficult or impossibile (as for paraplegic and tetraplegic patients), the interaction between human and machines is typically possible using the brain-machine interfaces (BMIs) by which, using specific patterns of neural activity, a person is able to send a command to an external device. In order to perform a correct utilization of BMI, users should undergo a training period during which their electroencephalogram data are acquired and a classifier is created, leading them to control an external device. Generally, most researchers neglect the importance of subject learning (e.g., both subject and decoder learn from each other) and they prefer to focus on machine learning techniques to increase decoder performances: as it is known from literature, the brain constantly modifies itself (neuroplasticity) and if a continuous re-calibration of decoders is performed, the subject could not learn enough to use BMI in daily life because he could not be able to keep his neural patterns stable; moreover despite they are considerably studied in the literature, BMIs are not yet developed enough to be processed and used by all, especially given the high mental fatigue and the high training period to maximize their performances. Therefore, it is necessary to set defined standards to improve training strategies and fully exploit the potential of individuals, thus allowing an extension of the use of BMIs outside the research field: to achieve these conditions, this thesis aims to demonstrate that different typologies of training can lead to increased performances in BMIs usage, showing at the same time how the neurophysiological patterns evolve during this period and how subjects get used to maintaining such parameters constant. In point of fact, the main intuition is based on the psychological evaluation of people and how

they manage to change their characteristics depending on habits: if the main objective will be to use, as experienced in this elaborate, a wheelchair, then those who immediately train with it, will be better and consistent in its use, compared to others who train in different ways.

To give to the reader a general view of this environment and to provide him the keys to a deeper understanding of the goal, this work has been divided in four principal areas: in the first one, it is given a brief recap of the brain and how it works with a detailed description of BMIs and the methods to create them; the second part is devoted to define the experimental protocol of the project, highlighting the different strategies chosen for the experiment and the importance of mutual learning during all the days of the trials; in the third part, there will be explained the methodologies of data processing and the algorithm used for the classification; finally, in the last chapter, all the acquired data are evaluated with multiple statistical and numeral approaches in order to demonstrate the differences and the advantages of a training typology respect to the other one, including also informations that are generally not covered in the literature. Indeed, the choice of the best training protocol can increase significantly both subject and decoder performances, decreasing the days required to master the usage of BMI and simplifying the approach: if a new standard and easier learning protocol is defined, then this technology could also be used by people with motor impairment, whose features are usually not stable, without the need of an hard work. In conclusion, this thesis demonstrates the effectiveness of these training sessions focused on end use of an external device and the role that different variables have in the evaluation of the neural correlates, especially in a mutual learning-based approach.

# ACKNOWLEDGEMENTS

I want to start by expressing my deep thanks to my thesis supervisor, Prof. Luca Tonin, for his unwavering support of my thesis research and study, as well as for his patience, inspiration, excitement, and vast knowledge. His advice was helpful to me throughout the whole research and thesis-writing process, especially for making the research environment welcoming and stimulating. Additionally, I also want to thank the committee and the University of Padua for letting me pursue a program of study that matched my interests and my objectives. My sincerely gratitude to my friends and colleagues, Marco, Gabriele and Federico: working as a team not only builds stronger bonds and more valuable connections, but it makes it easier and more fun to work. I am also thankful to all of the partecipants in my experiment for their time and effort in it.

I must express my heartfelt thanks to Nene for her unwavering support and encouragement in every single occasion and during our years of education. Thank you for enduring and supporting me at both good and bad times! A huge thank you also to Nik, CC, little Cloe, Noemi, Edoardo and Michele for coming, despite the times and distances, to my graduation and making this day even more special. Moreover, I would also like to thank those who are not here: thanks to Matteo and Noemi for the beautiful evenings spent in the canteen and with the board games, thanks to Erikitty, Fabiokitty, Chia, Vero and Turtle for the three years spent in happiness in Cagliari.

To conclude with fervor, my last big thank you goes to Diddy, my aunt Gianna and Cleto for being also present here but it goes especially to my mom for believing in me in all these years and helping me grow as a person.

# LIST OF ACRONYMS

**HMI** Human-machine interfaces

**BMI** Brain-machine interfaces

**ML** Mutual learning

**CNS** Central nervous system

**PNS** Peripheral nervous system

**EEG** Electroencephalogram

**MEG** Magnetic encephalogram

**fMRI** functional Magnetic resonance Imaging

**PET** Positron emission tomography

**CAT** Computer tomography

**ECoG** Electrocorticography

**fNIRS** functional Near-infrared Spectroscopy

**ALS** Amyotrophic lateral sclerosis

**EP** Evoked potentials

**ErrP** Error related potentials

**MI** Motor imagery

**FFT** Fast Fourier transform

**PSD** Power spectral density

**FS** Fisher score

**SA** Sample accuracy

**RD** Riemann distance

**BC** Between classes

**WC** Within classes

**AF** Accumulation framework

# Contents

# LIST OF FIGURES

# Chapter 1

# Introduction

With the evolution of brain analysis technologies, there has been a growing interest in trying to understand how this organ works, how it can be ”repaired” and how it can be used to facilitate people's lives and perform actions deemed difficult. In order to understand how the information obtained from the brain can be used, it is necessary to initially study its macroscopic and microscopic anatomical structure to make sense of what is being measured. In this first chapter we will then make a brief anatomical and functional reference, with particular emphasis on the characteristics most related to the experiment of this elaborate; we will then go to a review of current brain imaging technologies, their effectiveness in the diagnosis of specific diseases and their use in brain-machine interfaces. Finally, we will analyze the concept of Mutual learning (ML), how it is treated in the literature and how it can be a fundamental hinge for future developments.

## 1.1 Brief recap of structure and functionality on the brain

Every function of the human body, starting from the interactions between organs and with the surrounding environment, is managed and analyzed by the brain: processing the information in the form of electrical impulses, it is able to develop a feedback that, through the spinal cord, is sent to the affected body districts. This work is carried out by the Central Nervous System (CNS, composed of the brain and spinal cord) in close communication and collaboration with the Peripheral Nervous System (PNS, group of nerve ganglia and nerves outside the brain).

The brain consists of gray matter (neural soma) and white matter (neural axons) and is divided into the right hemisphere and the left hemisphere [1]; these zones are connected to each other through the corpus callosum and, although macroscopically similar, they have microscop-

ically clear differences in structure and, consequently, in their functions. The left hemisphere mainly controls the voluntary movements of the right part of the body, the ability of language, logical reasoning, calculation and thinking; In contrast, the right hemisphere manages the left part of the body and deals with the recognition of space and objects, imagination and intuition. In general, the two hemispheres can be considered as two functionally antipodes: one governs the pragmatic part, the other the abstract. A further subdivision, based on the location of the various districts (*Figure 1.1*), sees each hemisphere divided into frontal lobe, parietal lobe, occipital lobe and temporal lobe; each of these zones deals with particular functions, although none of them ever works as isolated and, even if only one of them is damaged, the whole brain functioning is compromised.

Located in the anterior position of the cerebral cortex, the frontal lobe is the largest lobe in which most of the functional areas are concentrated, especially those dedicated to movement: in fact, here lies primary motor cortex, aimed at the actual control of voluntary movements, and the premotor cortex that supports the previous one. In addition, the areas of spoken and written language control and the emotional control centre also reside here.

The temporal lobe is located laterally, protected by the temporal bone, along with the amygdala and hippocampus. It manages the processing of sound signals, coming from the outside environment, through the primary and secondary auditory cortex and it deals with the creation of visual memory and language; moreover, being the lobe dedicated to long-term memory, is often affected by pathologies of seniority or schizophrenic disorders.

In the parietal lobe lies the somatosensory cortical area, mainly responsible for the processing of tactile stimuli (feeling of pain, cold, roughness...); moreover, through the superior parietal lobe, the spatial orientation is managed while the inferior parietal lobe deals with the interpretation of the language.

Finally, the occipital lobe, the smallest of the four, guarantees through the primary and secondary visual cortex the reception and processing of all the visual information acquired and sent by the eyes and the optic nerve. This information is then sent to the parietal lobe and temporal lobe for further refinement of their decoding.

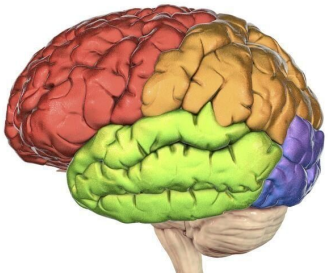


Figure 1.1: Lateral 3D view of left hemisphere with its lobes. The frontal lobe is coloured in red, the temporal in green, the occipital in purple and the parietal in orange

The information is exchanged and processed through electrical stimuli that are generated and spread in particular cells defined neurons. These nerve cells are divided into three main areas: the soma, generally of spherical shape, contains the nucleus and all the corpuscles necessary for its survival and organization; the axon is the extension towards which the electric stimulus (action potential) passes and propagates: it is regenerated thanks to the Ranvier nodes (i.e. areas where the myelin sheath, the membrane that protects and electrically insulates the axon, is absent) and then it spread throughout the body and reach the synaptic buttons through which, by means of neurotransmitter vesicles, the signal passes to the dendrites (linking sections between neurons) of the next neuron, along with other signals from different areas [2]. Depending on the anatomical district, neurons differ in shape, structure and function: sensory neurons transduce and send signals from outside (sensory somatic neurons) or from inside (sensory neurons visceral) to the CNS; motor neurons, on the other hand, receive signals directly from the CNS and send them to peripheral organs such as skeletal muscles (somatic motor neurons) or smooth muscles (visceral effector neurons); finally, associative neurons (interneurons) are those intended for the reception and organization of external stimuli, coordinating the respective exit information and reside internally at the CNS [3].

## 1.2 Signal acquisition technologies

To obtain information about the functioning and health status of the brain, particular techniques are needed to identify specific patterns required for the end purpose. As with any biological measurement, it is necessary to differentiate invasive techniques from non-invasive techniques:

the first ones, which involve the use of instrumentation that comes into contact with internal areas of the body, are able to provide more accurate information but at the same time can cause tissue damage and other problems; instead, the latter ones, are used in contact with the outer opening of a body area ans allow an higher level of safety for the patient but more noisy data.

### 1.2.1 Electroencephalography

Used for the first time in 1929 by Hans Berger, it is the most used non-invasive system for the acquisition of brain signals in both medical and research fields and it is based on potential differences detectable directly on the scalp [4]. In the outer part of the cortex, perpendicularly to it, there are pyramidal cells, particular types of neurons that depolarize or polarize depending on brain metabolism. The measurement of this activity takes place through electrodes, generally made by gold, placed on the scalp surface: due to the macroscopic size of the instruments compared to pyramidal cells, it will be possible to measure only the result of a collective activity (temporal and spatial summation) of the neurons of the examined area, which will depend on their degree of synchrony or asynchrony.

There are different types of equipment for EEG measurements (single electrodes, helmet...) but most involve the use of particular electrolytic solutions to be distributed between the electrode and the head, in order to maximize the conduction and reduce the skin-electrode impedance. Depending on the study carried out, a different number of electrodes are used: generally they are power of 2 (16-32-64-128-256 electrodes) and their position depends on the standard chosen [5]; as the number increases, the sensitivity in obtaining data of smaller neural populations increases too, but with a consequent increase in the computational weight of the data and increased preparation time and dressing of the headset. The acquisition of potentials takes place by differential measurement, calculated directly at the level of the analog structure, between the electrodes and the ground, an electrode that is ideally fixed at potential 0 and is usually placed in the forehead or one of the ear lobes. The data then undergo further processing both at the analog and digital level depending on what features you want to obtain. Regarding the latter, of particular clinical importance is not primarily the temporal evolution of signals, but more prop-

erly their behavior in the frequency domain. It is possible to identify four main frequencies, distinctive of certain situations of brain activity: the $\delta$ rhythm (0.5-4 Hz, 150 µv) is typical of deep sleep, childhood or some brain diseases; the $\theta$ rhythm (4-7.5 Hz, 75 µv) appears in newborns and in case of strong emotional tensions; the $\alpha$ rhythm (8-13 Hz, 30 µv) is recorded in an awake subject with closed eyes (in a state of mental rest), mainly in the occipital and parietal electrodes; the $\beta$ rhythm (13.5-30 Hz, 18 µv) is the most dominant in subjects awake and engaged in a brain task; finally, the $\gamma$ rhythm (30-42 Hz), appears infrequently and is characterized by very deep states of attention (please see [6]).

As already mentioned, EEG signals (*Figure 1.2*) are particularly noisy and difficult to interpret for more in-depth studies, and therefore they require various processes before useful information can be obtained; moreover, this techniques do not provide any information at the level of anatomical structure but only of functionality, and the correct measurement is not 100% guaranteed due to problems related to instrumentation, artifacts (which can be removed without unduly distorting the information content) or at the bad contact between skin and electrode (for example, the absence or scarcity of an electrolytic gel could lead to an increase in the electrode-skin impedance, therefore originating totally wrong signals in amplitudes and trends).

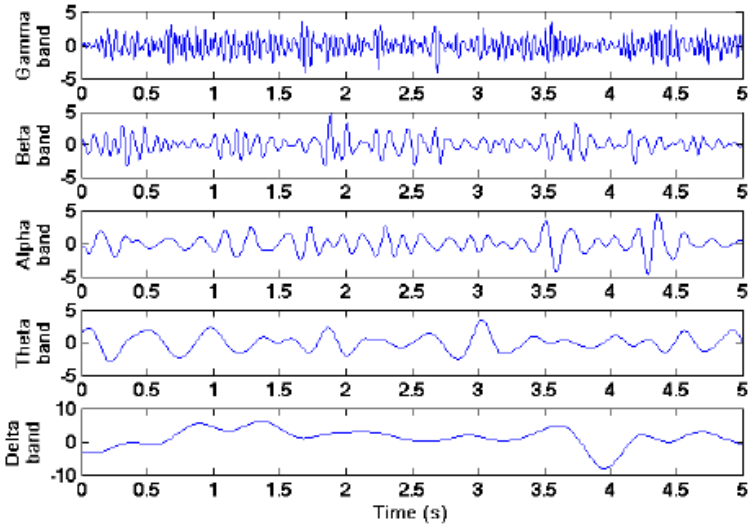



Figure 1.2: Time representation of EEG signal filtered in its different frequency-components. The amplitudes are expressed in $\mu$V.

### 1.2.2 Magnetoencephalography

Magnetoencephalography (MEG) is an other non-invasive technique [7] of neuroimaging that, unlike the previous one, exploits the magnetic fields generated by changes in electrical activity in the cortical area of the brain. These magnetic fields are based on the hypothesis that the neural families behave as dipoles perpendicular to the cortex and, then, they are able to generate an outgoing magnetic field: this measurement appears very advantageous in the study of high frequencies (unlike the EEG where they are very small in amplitude) and, in reference, in the study of epilepsy, in which the variations in activity are very strong, even if the machinery has big dimensions (*Figure 1.3*). In order to have a more complete view and higher spatial and temporal resolutions, the two technologies are usually used simultaneously [8] with the necessary shielding, in order to minimize the presence of electrical (EEG) and magnetic (MEG) artifacts .

Figure 1.3: MEG structure is set around subject's head: to avoid artifact, the patient should not move his head or do any type of movement of facial muscles. The acquisition is painless and, usually, used with a shielded EEG headset

### 1.2.3 functional Magnetic resonance imaging

Like the classical magnetic resonance imaging technique, the functional Magnetic resonance imaging (fMRI) is based on the principle of brain district response to an external magnetic field, adding a differential component over time that can identify voxels, whose value changes

depending on the oxygenation level (BOLD, Blood Oxygen Level Dependent) of the studied area [9]: hemoglobin is diamagnetic if oxygenated but paramagnetic if it is not, and this level can be detected and developed as a visual signal. Since the two technologies are based on the same machine, they can be used simultaneously allowing the evaluation of both anatomical and functional information (*Figure 1.4*). An higher level of oxygenation indicates an increase in neural metabolism and, therefore, shows the activation of the surrounding area compared to the baseline conditions. The scanned images have an high spatial resolution (it reaches a size of more than 1024x1024 pixels) but a reduced temporal resolution due to the acquisition time dependent on the parameters chosen by the specialist (T1-T2) and the limits of the machinery; moreover, it is not possible to determine, in the case of multiple BOLD signals, whether there is a correlation between them (especially at what time it started) and whether such signals originate from a change in functional metabolism or from external factors such as the use of excipients or drugs.

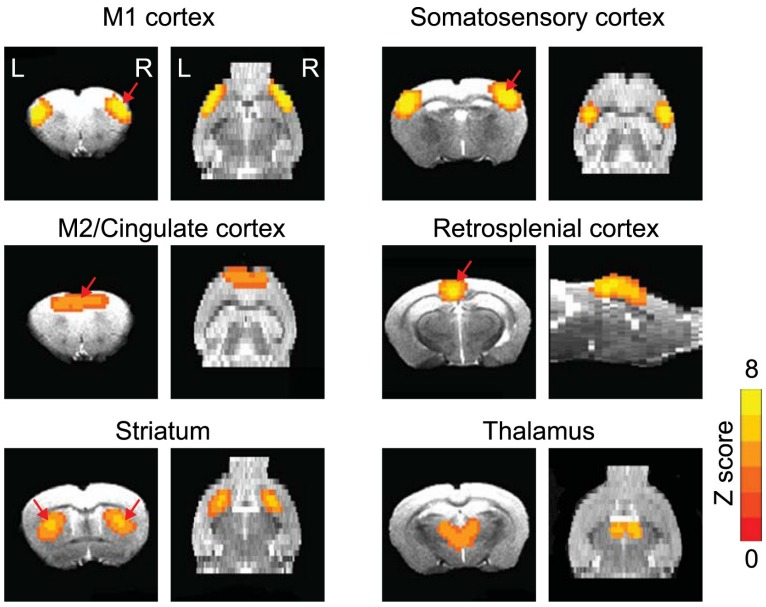



Figure 1.4: fMRI can used to estimate the brain metabolism along different planes, highlighting regions that can not be acquired with other technologies. Moreover, the anatomical information is preserved, making it easier to study more deeply more areas

### 1.2.4 Positron emission tomography

Positron emission tomography (PET) is a diagnostic technique of nuclear medicine [10] that uses labelled substances with radioactive markers to assess if there are problems in specific brain

districts; this evaluation depends on the injected substance (for example, glucose is used to look for cancer areas) and its body kinetics (the time it takes to reach the desired area), making PET a very versatile but extremely slow technology, with average times that exceed the hour. The construction of the image (*Figure 1.5*) takes place thanks to sensors sensitive to the emission of positrons following the decay of radioactive markers, and provides an anatomical-functional vision: the marker will decay when its substrate binds to a specific receptor, enabling to quantify not only its bond level but also the shape of the arrival structure. Like the computerized axial tomography (CAT) scan, the presence of radioactive elements can be harmful to the health of patients in the long term, and there are many restrictions on its use, especially in research field.

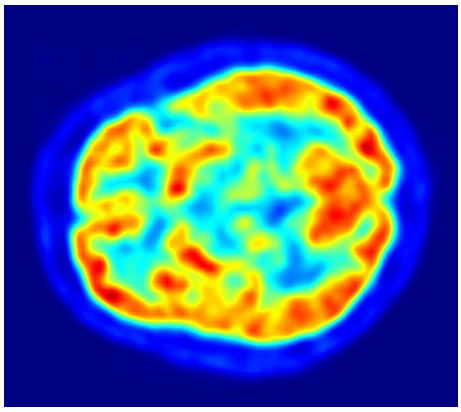


Figure 1.5: Transaxial slice of a brain acquired in 20 minutes: red areas represent high bound-level of the tracer while blue areas are the one in which there is no enough radioactive decay

### 1.2.5 Electrocorticography

Just like the EEG, the Electrocorticography (ECoG) has the aim of study the neural currents that develop as a result of brain activity too; however, it is an highly invasive technology [11], which involves the insertion of needle-like electrodes (*Figure 1.6*) directly inside the cerebral cortex. The electrodes used are significantly smaller than those of the EEG and therefore allow to have a very high spatial resolution ( 0.35 mm) and map individual neural families with greater precision, especially due to the reduced contact impedance; nevertheless, its use is mainly linked to the field of neurosurgery where it is necessary to ensure that there is no damage to brain tissues: moreover, because it is an invasive measure, there could be a very high risk of infection and, so, it can not be used frequently.

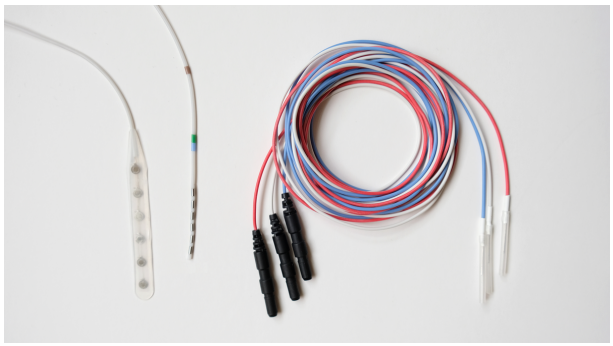


Figure 1.6: Example of needles used in ECoG measurements. The entire structure penetrates into the surface of the cortex but only the tip is conductive

### 1.2.6 functional Near-Infrared Spectroscopy

Similar to fMRI, the functional near-infrared spectroscopy (fNIRS) [12] captures hemoglobin levels in the cerebral arteries as a result of changes in brain metabolism through the evolution of light absorption (such as the pulse oximeter). The different corpuscles that make up the blood have different level of absorption and diffusion of light in specific wavelengths: hemoglobin changes its behavior depending on whether it is charged or discharged with oxygen, varying values within the near infrared spectrum (700 nm - 1mm); the increase in brain activity is related to a greater supply of oxygenated blood and, therefore, increased absorption in the wavelengths of oxyhemoglobin compared to a resting situation (values calculated according to Lambert-Beer law). The instrumentation used, even if it is extremely slow like the fMRI, can be very compact and portable (*Figure 1.7*), becoming a perfect candidate for daily use in the field of BMIs, as reviewed in [13].

Figure 1.7: The device is extremely compact and portable. The one above is able to acquire data from the frontal lobe

## 1.3 Brain-machine interfaces

Some of the previous techniques (EEG, ECoG, fMRI), beyond to purely monitoring and analyzing the pathologies, can be actively used to allow patients with physical disabilities (i.e., paraplegics and post-stroke patients) [14] to interact with the outside world only through the use of their still functioning areas through the human-machine interfaces (HMI); in particular, if the data are acquired according to the techniques mentioned before, then we will talk about brain interaction or BMIs [15]. These technologies have undergone considerable evolution in recent years, mainly due to new discoveries on the functioning of the brain and the increasingly precise acquisition of biological signals. Generally it is possible to divide the BMI in two large categories depending on the type of neural pattern that are analyzed: we will talk about exogenous BMIs if the decoding is based on a response to external feedback (EP, evoked potential and ErrP, error related potential) and endogenous if it does not depend on external events. Both types have both positive and negative sides and have a significant component of subjectivity, as each individual responds differently according to his health conditions, stress and fatigue.

### 1.3.1 Exogenous brain-machine interfaces

Exogenous BMIs exploit the responses of the brain following an impulse coming from the external environment, like an intermittent light or a continuous click: these responses are identified with the name of EP (also used for the diagnosis of neuro-degenerative diseases such as those described above) and they are related to different senses, that can be mainly visual [16], auditory [17] and somatosensory [18]. Depending on the type that is used, the useful signals are extracted from the brain areas corresponding to the task (for example, the data connected to visual information are analyzed starting from the electrodes placed in the occipital lobe) and are processed to derive a characteristic waveform: in fact, the EP are not immediately visible and recognizable by EEG signals but require particular types of extraction, more or less meticulous, depending on the purpose of the analysis.

Generally, there are specific features that allow to identify the presence of EP and they can be obtained from the grand average, that is an averaged signal obtained by the mean of several

window synchronized according to the start of the stimulus, under the assumption that the EEG is affected by a white noise and, therefore, with mean equal to 0: it is evident that such analysis can provide generic characteristics and it can be highly problematic in case of errors in the choice of signals or time synchronization.

In the literature there are proposed several exogenous BMI based on different typologies of EP: here, we talk about the P300 [19] and Errp [20], both visible along the central line of the frontal lobe, appearing when the state of attention of the subject changes following an external event. When the user notices something, a wave mainly characterized by a localized peak about 300 ms after receiving the stimulus is created (a considerable variation of this value may indicate a pathological state such as the onset of ALS, Amyotrophic lateral sclerosis): this EP is mainly used, together with ophthalmology, in the BCI Speller [21], that is a particular type of HMI that allow verbal communication thanks to a screen containing the letters of the alphabet (*Figure 1.8*); this technology can be very useful for those persons that, due to pathologies or trauma, are not able to talk properly. The Errp instead, originates following the recognition of an error (personal or other) and its use is still under study because it is not easily recognizable in all subjects.

Being based only on brain feedback to external stimuli, exogenous BMI can be used with a minimum training, allowing its use even in a short time; on the other hand, the user could get used to the stimuli and get tired over time, leading to a progressive reduction and disappearance of the EP, making it extremely difficult to use succefully the BMI.

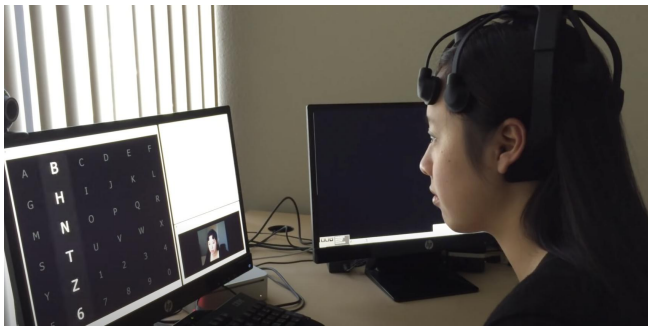


Figure 1.8: The subject is able to choose one letter among these that are showed in the monitor: the choice can occurs in a very short time period with reduced false positive.

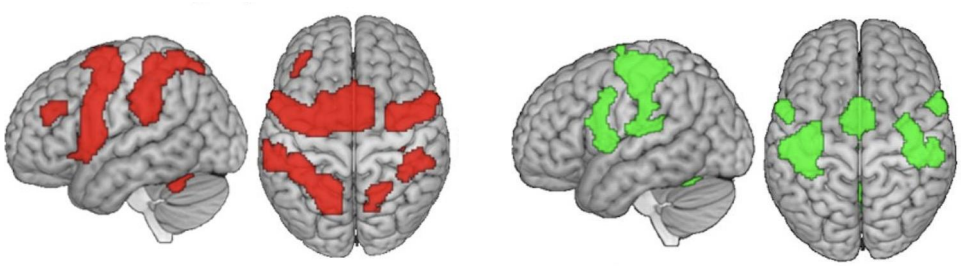
### 1.3.2 Endogenous BMIs

In contrast to the previous type, endogenous BMIs are based on self-modulation (or modulation of brain activity in the absence of external stimuli) of certain neural areas. According to the literature, the most commonly used type of endogenous BMI (*Figure 1.10*) is based on motor imagery (MI) [22], that is the ability of an individual to have a kinesthetic imagination of a movement of his personal body part (typically hands and/or feet) and to produce a modulation in the motor cortex, similar to what he would have with the real movement. MI BMIs employ the user's endogenous brain activity in the absence of any external stimuli: When imagination occurs, there is a progressive de-synchronization of neural activity in different areas (depending on the body part connected to the movement) in the center-frontal area of the cortex and then a returning to the initial situation when the activity is finished. The imagination and the execution are similar in terms of brain activity, even if there are some areas that are involved only in MI [23]: the spatial localization is mainly on the motor and sensory cortex (*Figure 1.9*). It is particular interesting the way in which the neural networks change during the kinesthetic imagination: although the main motor cortex was not consistently proven to be active [24], the MI network contains various areas that are recognized to be important during real motor execution. The type of MI tasks utilized, the modality of MI, and the body part engaged in the movements all appear to have an impact on how consistently the overall MI network is activated. When someone pictures the movement of his limbs, a significant fronto-parietal network is engaged and also the subcortical and cerebellar areas are consistently activated by MI. Moreover, it is shown that the time in which this network change is similar for real and imagination movements, proving that the required computational ability of the brain is approximately the same. Finally, the activation of these areas can drastically change if there are dehabilitant pathologies: among people who are affected by Parkinson's disease ([25]), there was greater difficulty in completing MI tasks compared to health subjects due to damages on motor areas.

With a more in-depth analysis of these zones, using an acquisition system like an EEG, it is possible to notice changes in the frequency behaviour according to the cognitive process: by analysing the baseline (i.e., in a rest situation) of the logarithmic spectral power, all the

information can be extracted from the attenuation in $\alpha$ and $\beta$ bands that occurs during the imagination task. In the areas in which MI occurs, the particular rhythm that describe better the change of neuronal activity is the $\mu$ that has a frequency range between 7.5 Hz and 11.5 Hz and its suppression is a typical sign of kinesthetic imagination.

To facilitate the subject's mental effort of imagination, it is recommended that the anatomical parts thought in motion were perceived in tension and, above all, perceived as their own (not enough to think of "a hand" or "a foot").



(a) Areas dedicated to motor imagery

(b) Areas dedicated to movement execution

Figure 1.9: MI task involves larger regions than movement execution due the activation of areas dedicated to the imagination.

The conversion of the user's intention into a mental image of a motor action is a fundamental idea in these BMIs and provides an interface for communicating the user's purpose without limb movement (even if, for people with stumps, the movement is allowed). One of the most study field in which MI BMIs could be useful is in neuro-rehabilitation, and studies have shown that repeating MI may be beneficial in this type of treatment [26]. The MI BMI paradigm includes training (calibration phase) and testing (evaluation phase) sessions, similar to other BMI paradigms: during the first one, a classifier is created as a result of feature selection and extraction while, in the second one, the user is able to control the BMI according to his features.

Another type of BMI, which no longer involves the use of an EEG but directly of an MRI, bases the operation on BOLD signals [27]: as explained in the previous paragraph, activation of a zone involves a change in glucose metabolism and a consequent increased need for oxygen, which alters the magnetic field allowing so the detection by the machinery. Unlike the first one, it is evident that the latter one, despite being able to more accurately detect changes in activity

along the entire cortex, still does not allow an everyday contexts usage due to the machinery dimension, the difficulty to manipulate and its slowness in data acquisition.

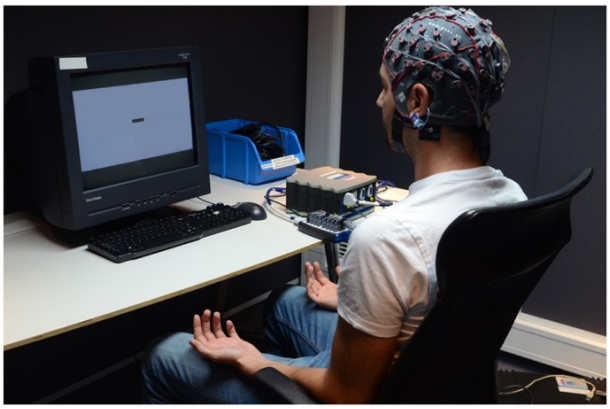


Figure 1.10: The subject is modulating his motor-cortex neurons by the imagination of limbs movement, without doing the real movement

The main point in favor of these systems is their high versatility and long-term use: not depending on any external event, there is no risk of habit (resulting in a reduction of the appreciable response) and above all, there is the possibility to be used when desired, because the user will decide when and how to activate them only with the use of his imagination [28]. This, however, involves a considerable mental effort that not everyone can successfully carry on, so a continuous training is required to maximize the capacity of usage that can last even months; moreover, unlike exogenous BMIs that require information from few or even one channel to work, MI BMIs need signals from an entire area (sensory-motor cortex): therefore a multi-channel recording is needed in order to have clean and usable results. Finally, the current self-paced BMI have also a reduced number of degrees of freedom, or the different actions that can be carried out with different neural modulations: this constraint makes extremely difficult to use complex technologies in three-dimensional spaces using only endogenous BMI (in fact, they are always supported by other types as showed in [29]).

### 1.3.3 Limitations and uses

At current stage, due to the limitations mentioned above, more complex BMIs are mainly used only in research field: today's non-invasive capture technologies do not yet provide clean signals

suitable for detecting more precise changes, especially for the attenuation that is found at the level of the scalp, which involves a significant loss of information. At the same time, invasive (using techniques such as ECoG) or partially invasive (Neuralink [30], (*Figure 1.11*)) BMIs can produce very strong signals from small brain regions, which allows the reduction of processing time and the creation of finer commands. Of course, because they can cause infections or damage to the brain, because of their high cost and also because of the long application time, they are used only in very narrow areas and only under the supervision of a medical team, unlike the non-invasive that can be used by everyone at any time. In addition to device control, BMI could been particularly effective in rehabilitation therapies for post-stroke patients, especially when connected to robotic devices that help with limb movement.

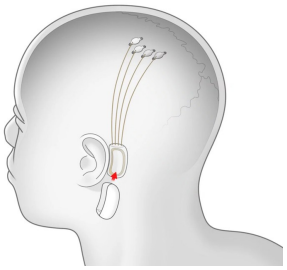


Figure 1.11: First version of invasive BCI by Elon Musk. The electrodes are placed directly on the cortex and the are fixed with little needle; the information is acquired and transferred to computers with bluetooth technology.

## 1.4 The concept of Mutual Learning

Most of BMIs that provide to users the control of an external device, are based on closed loop (proposed, for example, also in [31]): the interaction between subjects and computer doesn't work in a single way but in a request-response mechanism. In fact, when the partecipant sends a command, the device give him a feedback which can be of various types (mechanical, visual, auditory, electrical) in order to return the information about the accomplished task: this technique allows the user to understand how the response changes according to what he does, leading him to engage more and understand what are the best methodologies to maximize his intentions and reducing the effort to perform them. Generally, for classic endogenous BMIs, the feedback generated is visual and it occurs on the last part of the loop (*Figure 1.12*): after the acquisition of subject's signals, the data are processed in order to extrapolate the best features

(information that describes more the characteristics of a signal as highly discriminating and that can be used for its identification) for the classification and the delivering of the command to the external device. Moreover, the feedback can be originated from other sources outside the loop too, like proximity sensor in case of BMI-driven wheelchairs.

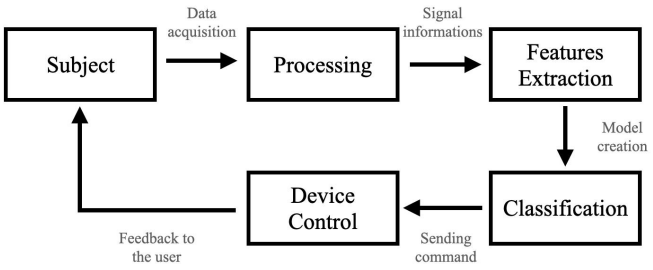



Figure 1.12: Schematic view of BMI closed loop. By studying the response of brain according to the feedback received, is possible to increase the users performances.

To achieve good results, it is necessary that every feedback must be generated in the shortest possible time after the sending of the command, in order to maximize the user's awareness of what is happening and allow him to correct possible errors.

Precisely for this reason, training is an essential part for a proper usage; above all, it is extremely necessary to build the entire algorithm carefully, studying the subjective characteristics of each user and evaluating the most important features to be analyzed: from user's responses to feedback, it is possible to study the evolution of his neural patterns and improve the performances for his subsequent BMI usages.

Most researchers in the literature mainly focus on identifying methods to improve BMI algorithms, with a focus on deep learning and machine learning and new EEG signal features, together with newer technologies. Moreover, it is particularly useful to focus on the training and learning methods of the subjects, evaluating new strategies and maximizing the performances of BMIs. A relevant approach is that of Mutual learning, used and described in [32], based on the characteristic of the three pillars of BMI (subject, machine, application layer/algorithm) to learn from each other (*Figure 1.13*): in fact, researchers try to maximize the algorithm in order to fully represent the needs of the user, without allowing the user to learn how the system works and to adapt accordingly. This adaptation is very important because, as well as increase

the chances of correct use of BMIs and their performances, it allows to drastically reduce the number of re-calibrations of the classifier, or the moments when the latter must be corrected because it can no longer decode the intentions of the subject. With a stable model, it is possible to quantify more the degree of training of the user and, especially, it allows the creation of a stable system over time and not dependent on unwanted fluctuations due, for example, to the user's mood. Therefore, the evaluation of neuro-plasticity associated with this new type of protocol, would allow a better and immediate use of BMI in everyday life even for subjects with motor disabilities and in more difficult situations, reducing stress and training time and allowing them to improve their quality of life.

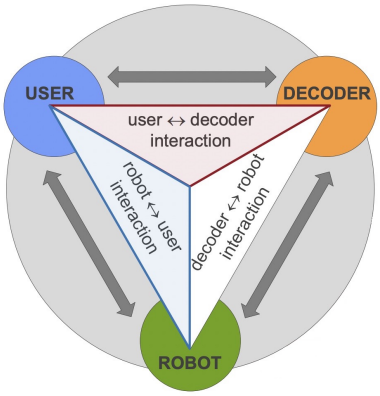



Figure 1.13: Representation of the ideal connection among the three pillars of BMI. The output of the BMI can also be corrected by other external devices like proximity sensor

# Chapter 2

## Motivations and Objectives

As highlighted in the introduction, currently, the use of BMIs is not very common in the areas outside of research, fact penalized especially by the long training times needed to ensure proper use. Moreover, the lack of a uniform protocol and a standard of operation of the algorithms, makes the transition from one use of a device to another particularly complex, precisely because of the absence of common guidelines that allow a cross-use or reduced adaptation time. For these reasons, this thesis sets as its main objective the evaluation of the benefits that a training based on ML can offer and above all, to quantify and demonstrate that the performance of BMIs do not depend exclusively on machine learning algorithms, but are directly related to the sensations, objectives and link with the machinery used by the user. According to this last point, in this experiment will be evaluated the consequence of a more intense feedback (wheelchair rotation) respect to the classical visual effect, under the condition that an improvement of closed loop follows an increasing of the performances.

In addition, by reducing training times and improving the latter's characteristics, it would be possible to increase the use of this technology for different areas, from remote or dangerous workplace tele-operations, to rehabilitation for post-operative patients stroke or with debilitating conditions such as ALS: given the time required for traditional therapy, this technology could dramatically improve patients' lives without, in addition, the need to have long sessions or to create new models; this could be possible if new methodologies that aim to create a compromise between subject and algorithm are applied. Therefore, the question is if there is enough differences, in the third day of training, between people that trains in different ways.

# Chapter 3

# Methods

This chapter is dedicated to the presentation of the methodologies and the experimental setup in which the experiment was conducted: in the following paragraphs, particular attention will be paid to the composition of the groups of participants and to the data acquisition and processing, together with a timely description of the analytical metrics.

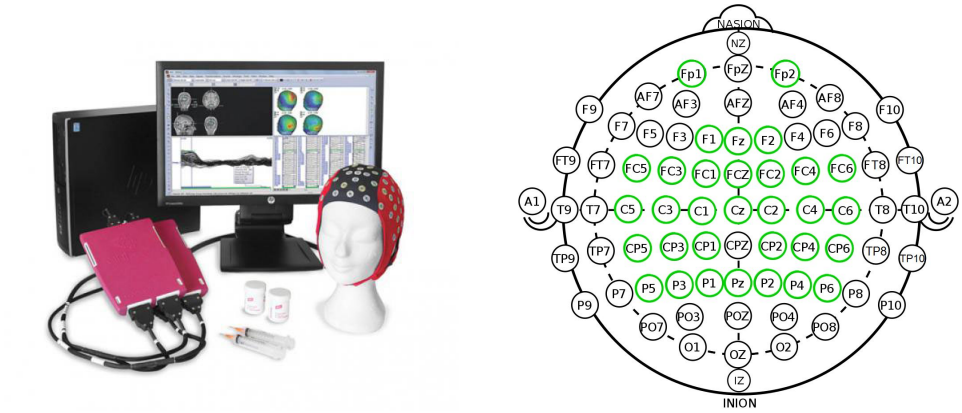
## 3.1 Partecipants

Eight subjects participated in this study (average age $25.5 \pm 1.5 years$, 4 males and 4 females) without any pathology and without any previous experience in the use of BMIs. Each of them was pseudo-randomly assigned to one of two groups: the first(group "on desk", a8-b9-b1-a5), carried out a training at BMI according to standard methodologies, with a visual feedback given by a monitor; the second (group "on wheelchair", c1-a3-c4-c6), instead, used since the first sessions a robotic wheelchair that rotated on itself depending on the output of BMI. All participants, before starting the first day of experiment, have viewed and signed an informed consent in which all information for the treatment of privacy and the use of recorded data were illustrated (for this purpose, each user has been anonymized to ensure greater security). All subjects were asked to be in the laboratory for a total of three discontinuous days.

## 3.2 Signal acquisition

Throughout the duration of the experiments, EEG signals were extracted and used to develop non-invasive BMI algorithms. To acquire the EEG signals, the *AntNeuro eego sport* helmet is used: although there was a total of 64 electrodes available, only 32 have been used for measurements, arranged according to the standard 10-20 placements (*Figure 3.1*). To reduce the

electrode-skin impedance, an electrolyte solution was applied to each electrode several times during the same session in case there was a strong presence of noisy signals; the reference electrode was Cz and the sampling frequency was 512 Hz. The EEG signals were then saved in standard .gdf format along with task labels and other information (such as the duration of a task, the beginning of the session) by using the Ros-Neuro system [33] [34] which would also deal with the online phases (i.e., the evaluation mentioned above).




(a) *AntNeuro eego sport* setup (b) Electrodes used

Figure 3.1: The headset is fully portable and all the data processing and storing is inside the pink amplifier. Among all the 64, only the most useful group of 32 electrodes (highlighted in green color) is chosen for the experiment

In order to evaluate the correct acquisition of signals, they have been plotted (after a filtering between 1 Hz and 40 Hz to reduce the noise and possible artifact) and the user was asked to perform three simple tasks, with the aim to show recognizable patterns to quantify the degree of cleanliness of the signal: the subject had to blink and, then, repeatedly move the jaw to highlight muscle artifacts first in the electrodes Fp1 and Fp2 and then in the surrounding area, that area affected by the movement of the facial muscles. Finally, to look more closely at the frequency components, the user had to close his eyes and relax as much as possible to highlight the $\alpha$ waves in the occipito-parietal area. If only one of these variables had not been displayed correctly, it would have gone to a better arrangement of the headset and gel.

## 3.3 Experimental protocol

The sessions of the experiment were distributed in three days, spaced on average of $3 \pm 2.58$ days from each other to allow individual subjects to have a period of pause between uses, thus allowing a better evaluation of their neuro-plasticity. Of these three days, the first two were dedicated to calibration (i.e., creation and subsequent improvement of the classifiers to be used in BMI) and the subjects underwent training and evaluation sessions with the methods dictated by the group (in front of a computer for the first group, directly in the wheelchair for the second) (*Figure 3.3*); during the third day it was not possible to change the classifier created in the previous days (with some exceptions that will be described in the final discussions and considerations) and both groups performed that session in the wheelchair (*Figure 3.2*). If some subjects failed to use the BMI in the first two days of calibration: among the metrics used to decide the transition to the evaluation phase were the stability of the subject's features over time, the speed at which it was able to perform the task, the percentage of correct tasks and the level of fatigue. Each session had a variable average duration of about two hours, a quantity strongly dependent on the fitting of the headset, the application of the gel, the evaluation of the cleaning of the signals and, especially, the time taken by the subjects to complete the evaluation (8 runs were completed every day, with some exceptions for subjects who controlled the BMI perfectly and for those who could not).

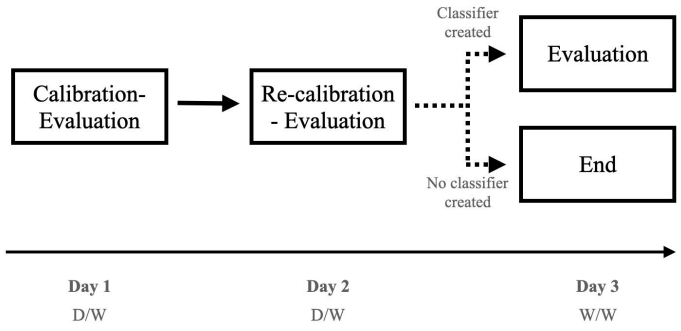
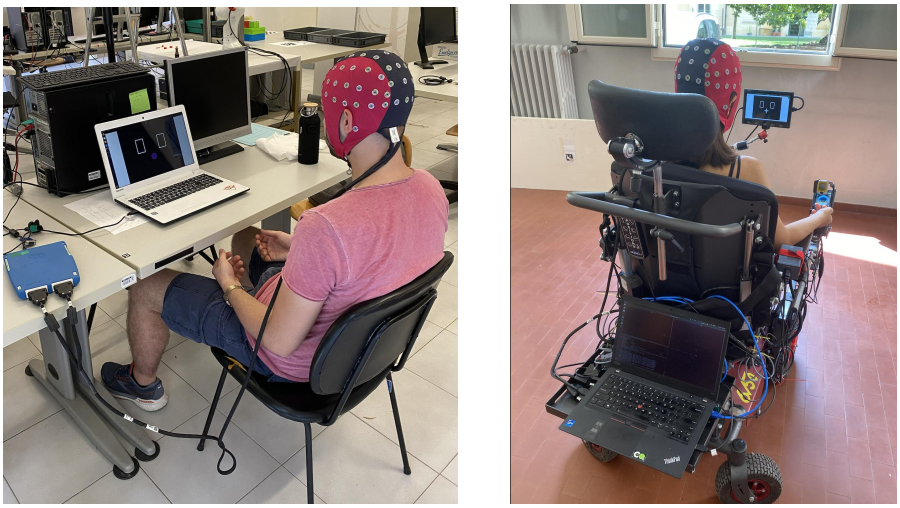



Figure 3.2: Schematic view of experimental setup. D means "session on desk" and W "session on wheelchair". This letter indicate the modality according to the first and the second group.

During the first day, we started with the creation of the classifier: the number of runs needed

varied from subject to subject but generally, three were enough to identify stable features that could guarantee good performances. Once the model was created, we moved on to an evaluation phase that, according to the metrics listed above, allowed us to understand if we could move on to the next day; in the case of the first day, if the BMI had not had the expected outcome, however, we passed to the second day building the sessions as if it were the first (in fact, the second day would be used mainly to check if the features of the subject had remained stable and, to the limit, to further refine the model).



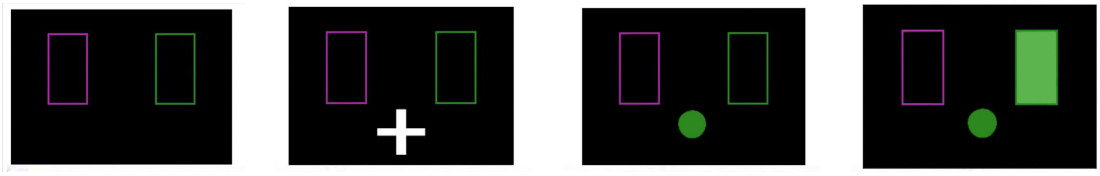
(a) Training on desk (b) Training on wheelchair

Figure 3.3: Two types of training according to the group. In both situation, subjects have to keep their hands and their feet as much relaxed as possible; the feedback on desk is just visual while the one on wheelchair is both visual and of movement

## 3.4 Motor Imagery and visual paradigm

The runs, divided between offline (calibration) and online (evaluation) were composed by 20 tasks, each depending on the type of kinesthetic imagination (10 for ”both-hands” and 10 for ”both-feet”, proposed in random order to minimize the risk of habit and therefore decrease in features) that the user should have performed: for the calibration session, the completion period was about *5 minutes* while for the evaluation it varied depending on the performance of the classifier and the subject. During the experiment, to the user was presented on the screen, alternately in the form of figures and colors, the task to be accomplished (*Figure 3.4*): the presence of a green/purple circle suggested the user to start imagining to move both feet/hands contin-

uously until the complete filling of the corresponding bar placed at the top. Before starting, it was proposed to the user to move the hands or feet rhythmically with an increasingly slow pace until the movement was completely absent but there was still the feeling of contraction and relaxation: because of the fundamental requirement of this experiment, that is not to have any previous experience with BMI, this exercise was fundamental to better prepare participants for what they should have done. In order to reduce as much as possible movement and ocular artifacts, it was required not to blink or move the head while filling the bars; moreover, before each cue (start of the task, *1 second*), a fixation cross (*1 second*) appeared in the centre of the screen, alerting the user to the start of the test and acting as a fixed point on which to look for greater concentration and less distraction due to external movements in the room, which still had to be as quiet as possible. If in offline runs the bars rose automatically (continuous feedback, *5 second*) correctly compared to the proposed cue, in online runs was the user to control them and therefore could rise at different speeds, up and down or, one could not climb at all despite the effort of the subject, accordingly to the output of the BMI decoder. In order not to decrease the level of concentration, between one task and another there was a period when the user could relax and prepare for the next cue: being randomly proposed, limb relaxation was mandatory to prevent the muscle contractions of one area from corrupting the acquired data for the other. In case of fatigue, subjects could make small breaks between the runs, especially to cool their eyes, but without touching or removing the EEG headset due to the presence of the gel (if different electrode areas had come into contact, a short circuit would have been created and the resulting signals would have been totally wrong). Finally, to facilitate the evaluation approach, two threshold parameters, one for the both-hand task and the other for both-feet, could be modified according to the needs: if in fact the user had difficulty in picking up the bars, these parameters were lowered and tested in the next run; conversely, if a bar rose too fast without being checked, then its threshold parameter was increased. It is clear that increasing or decreasing these values represents a highly subjective trade-off that had to be evaluated on a case-by-case basis: if the value had been too high, it would have taken more time to complete the task but there would have been more control by the user; vice versa, if the value had been too low, the bar would have risen very quickly escaping the real intention.



(a) Inter-stimulus interval (b) Fixation cross (c) Cue (d) Continuous feedback

Figure 3.4: Pipeline of visual paradigm proposed to users. After the inter-stimulus interval, a fixation cross is proposed and, then, the cue: now starts the imagination of movement (continous feedback) until the bar is filled. In this image the cue corresponds to the task "both-feet" with consequent rotation of the wheelchair to the right. This paradigm is equal for both groups.

## 3.5 Signal processing

As a first step for the creation of the classifier, EEG signals were filtered using two Butterworth filters (order 4 to speed up filtering for online applications) to minimize noisy and unnecessary components: in fact, the frequencies actually useful for the purpose of the experiment and corresponding to those modulated during motor imagery are those belonging to $\mu$ waves and the first $\beta$ waves.

### 3.5.1 Laplacian montage

During the first days it was necessary to create classifiers able to identify the discriminators between the two classes of BMI and, to do this, it was necessary to focus on the brain regions involved in the action, namely those of the motor-sensory cortex. According to the standard 10-20 and the headset used, we mainly refer to the front-central-parietal electrodes (Fz, FC3, FC1, FCz, FC2, FC4, C3, C1, Cz, C2, C4, CP1, CP2 ), that is those in which the differences between hands can be noticed most (mainly the areas of the electrodes C3-C4) and feet (zones of Cz). Being of their nature damped and noisy, it is necessary to increase the information content of the EEG signals by applying a spatial filter (that is a sharpening filter able to increase the signal-noise ratio and consequently the signal sharpness level) and highlight more the changes in field potential along the scalp surface. In the case of this experiment, a Laplacian filter was created and used [35] (matrix 32x32, size corresponding to the number of channels used) (*Figure 3.5*): this is a filter based on the second order differences of the electrode chosen with respect to its

surroundings for each of the individual samples of each temporal moment. The absence of this particular application would have caused the presence of too much noise that then would have poured on the spectral analyses, which will be described in the following paragraphs, making the model and the BMI practically unusable.

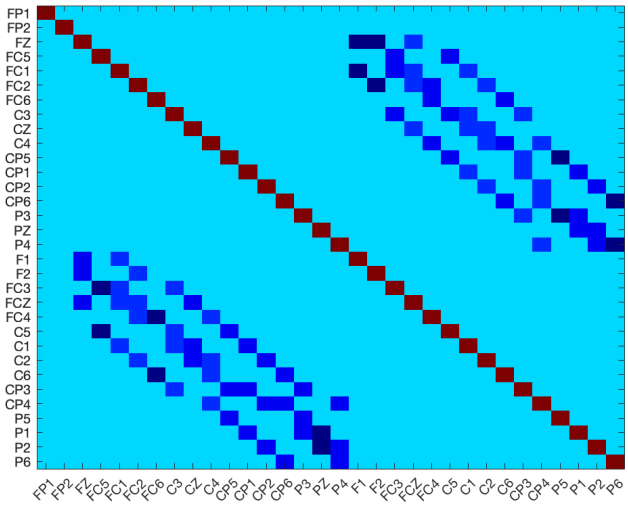



Figure 3.5: Colored image of diagonal Laplacian matrix. The main diagonal is referred to single electrodes while different blue tonalities shows the areas were spatial filtered is applied

### 3.5.2 Power spectral density

The analysis of signals directly in the time domain does not allow to identify features suitable to maximize the discrimination between the two tasks of MI BMI. According to what is proposed in the literature [36], cognitive processes related to the motor aspect not only generate responses similar to event related potentials (ERP), but present a particular response in the combination of neural work: by the analysis of the signal spectrum, it can be noticed that at the motor imagery, a desynchronization (ERD), that is a phase in which the cells no longer work in unison, producing a power change detectable especially in the entire alpha band with a sharp decrease in amplitudes with respect to the resting situation, of the neurons of the primary and somatosensory cortex is generated; when the task ends, it appears again a synchronization (ERS), with a consequent returning to the rest situation (*Figure 3.6*). However, this configuration is extremely variable among people and generally recognizable only if we consider the individual bands as a whole, mediated for all the frequencies of which they are composed: such situation can be

useful only in a case of study and not in a real use of BMI because most subjects using MI BMI do not modulate the entire band but only some sub-bands, unlike what described by ERS/ERD. To overcome this limit, methods have been studied to increase, as was the case for the spatial domain, the spectral resolution to have data linked as much as possible to multiple groups of frequencies and not to entire bands. Based on this, the variation of the power spectrum is analyzed (that describes, using discrete frequency component, the distribution of power that signal have in a specific time period) in various time blocks in order to obtain a spectrogram that links temporal information to spectral information.

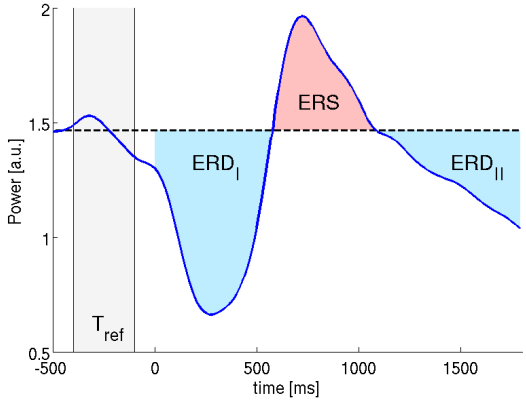



Figure 3.6: Temporal evolution of general ERD/ERS wave. When an imagination occurs (0 - 500 ms), the neurons start to fire alone generating less amplitude in the EEG signal (there is no more temporal summation, ERD) compared to the reference period. When there is a return to the rest state, neurons work again together (temporal summation, ERS)

Since it is pratical impossible to construct a perfect spectrum (since an infinite period of time is required in order to calculate the fast Fourier transform (FFT), we try to create an estimation as close to the correct version using methods of spectral mean as can be that of Welch [37] (the algorithm used for the creation of the features is based on this method and it generates the PSD, power spectral density). The signal is divided into a finite number of windows whose length (we chose, for best results, *1 second*) is chosen according to the requirements of the algorithm and such windows are further separated into overlapping segments for each shorter period of time (*62.5 ms*): for each of these, the spectrum is calculated by FFT for each desired frequency and then these values are then averaged representing the average value of the spectrum in that

signal portion; the same windows must be overlapped to maximize spectral resolution.

$$\mathcal{R}\left\{\frac{1}{2\pi}\int_0^{\infty} 2F(\omega)\mathrm{e}^{i\omega t}\,d\omega\right\} \tag{3.1}$$

At the end of the process, there will be a three-dimensional matrix (windows x channels x frequencies) containing all the spectrum information needed to evaluate, window by window, the mental state of the subject and to identify the start and end periods of the task: in addition to providing a significantly higher resolution than that of ERD/ERS, the Welch method, being based on the average of segments, allows to reduce the noisy component (if, even after preliminary filtrations, high-power components remain, the average with the rest of the segments would lessen their influence) and above all, make the spectral calculation faster than other methods such as Bartlett's [38] because the previous FFTs can be reused to calculate the following ones, allowing an online use of this algorithm.

Finally, a base-ten logarithmic transformation is applied to linearize the spectrum, to have a simpler view of its trend and to facilitate the creation of the model as will be described in the following paragraphs. Ultimately, therefore, a PSD-based approach allows not only to analyze more in detail the evolution of sub-bands (in this experiment was set a vector that starts from 2 Hz and reaches 48 Hz with resolution of 2 Hz) but also to clean up the spectra and have more precise information for the various periods of time, allowing to identify the individual windows of interest: once calculated, through a similar algorithm, the windows are separated according to the task (both hand, both feet and rest) thus moving from a three-dimensional matrix to a four-dimensional matrix that allows the identification of differences between the PSD, with the aim to create a classifier based on them.

### 3.5.3 Features extraction

In order to create a model useful for the control of the BMI, it is necessary to identify the features that discriminate in the most correct way the execution of a task respect to the other one: one of the objectives of this process is the analysis of the evolution of neural correlates

due to neuro-plasticity of subjects brain; moreover, it is necessary to reduce the over-fitting phenomenon and the computational heaviness of algorithms in the online usage by creating a classifier based on a limited number of features. As defined in the previous paragraph, the PSD calculated from the signals has a rather high dimensional and provides a very high number of usable data, most of which, however, are not discriminatory and, despite that, they could lead the model to make wrong predictions about the outcome at that time: there is therefore a need for statistical selection methods based mainly on the neuro-physiological notions listed in the literature [39]. Although the PSD is built on a very high frequency scale, only small subgroups of them are useful to decode the intention of movement, as it is characterized by changes in the spectral power in $\mu$ waves and in the first $\beta$ waves on specific areas [40], that are those typical of the motor cortex; known this, the number of the features is drastically reduced to few channel-frequency pairs: for this experiment six frequencies (according to the resolution chosen for the algorithm) and about 13 of the original 32 channels (corresponding to those in which the difference of PSD between one task and the other is more evident, information values greatly increased also by the laplacian applied in the first processing) will be considered.

Of course, each subject will have a different neural response from the others (as evidenced by [41]) and, therefore, this point must be taken into account in the calibration phase: this is the main reason why a single generic model can not be created. To identify the frequencies and channels most affected by the difference between the two tasks of the MI, it is possible to use different statistical approaches, even if the most used, especially for its application speed, is the Fisher Score (FS) [42]

$$FS(k) = \frac{abs(\mu_2(k) - \mu_1(k))}{\sqrt{(\sigma_2^2 + \sigma_1^2)}} \tag{3.2}$$

For each calibration run carried out, this value is calculated for every channel/frequency pair and a map is generated: in this case, it will have a size of 32 x 26 (channels x selected frequencies). This map provides a first visible result of the pairs that are most discriminating for the two tasks and that are represented by an higher coefficient than the others. Once the corresponding

matrices for the runs are available, it is necessary to obtain a single one through an average of the single pairs, making more evident those features that have remained constant during all the runs and reducing the discontinuous ones, in order to identify their stability, aim of this study (*Figure 3.7*). This statistical index is able to show the time evolution (run after run) of channel-frequency pairs in which the activity difference (PSD) between the both-hand task and the both-feet task is stronger: of course, these values are strongly dependent on current modulations and can change in intensity, despite their representation in the same way (same colors) in several maps.

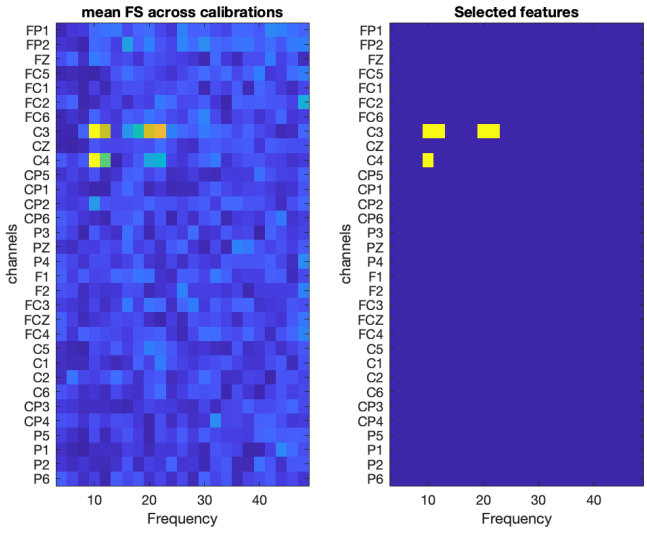
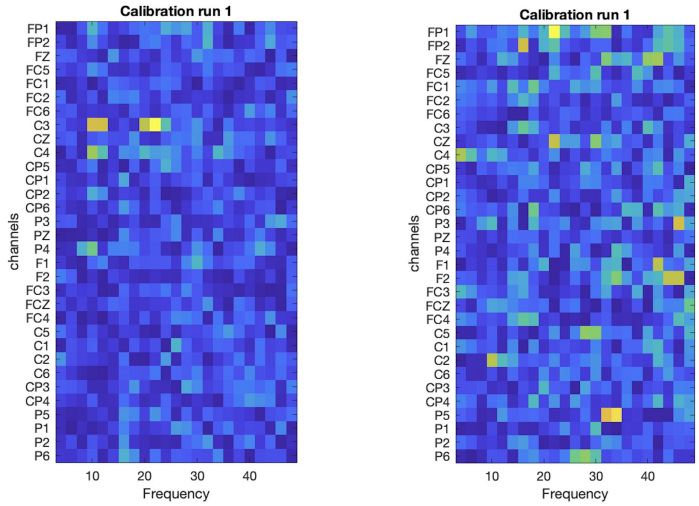



Figure 3.7: Example of Fisher maps generated by the averaging of every calibration runs. In the right there is the masked version where the chosen features are highlighted

### 3.5.4 Classification

Once obtained the average map of the FS between classes (its creation is not necessarily based on the use of successive runs or only offline, but varies according to the needs that the scholar considers most appropriate), we move on to the selection of the most significant features: the choice of them depends not only on the highest values found, but requires a thorough knowledge of the theory a priori. As explained above, only the central and front channels and only the $\mu$ frequencies and the first $\beta$ waves are typical of the MI; in case there are evident indices outside this group, they should not be taken into account because they originate from random situations outside the self modulation, as can be eye movements (acquired by the two electrodes Fp1 and

Fp2 placed on the front of the subject that play the role of the classic eog electrode present in many helmets for EEG analysis), distraction from external events, changes in the level of attention [43] or stress, and muscular artifacts (30-40 Hz) [44]. In addition, the features to be selected should be the same, or at least in part, for each run analyzed, both for a matter of study on the evolution of neuro-physiological patterns, both to have a model as stable as possible, not dependent on temporary fluctuations and with high performance, insurable only by continuous properties (*Figure 3.8*).



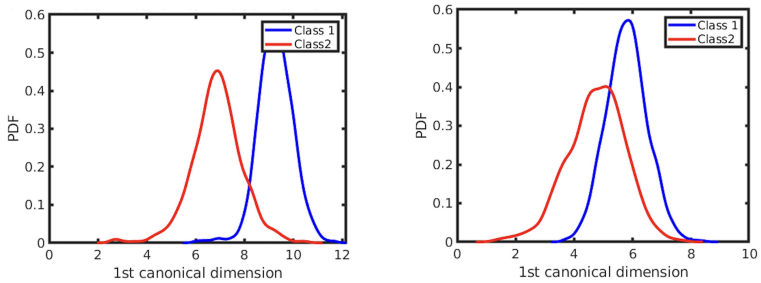

(a) Good features (b) Bad features

Figure 3.8: Figure *a* shows an example of correct features: the higher values are in the sensory-motor cortex electrodes and in the $\mu$ and first $\beta$ bands. Conversely, in the figure *b* the higher values are only in Fp1 and Fp2 electrodes: this can mean that the subject moved his eyes during the experiment. The first map can be used for classification unlike the second one.

Having ascertained the existence of all these requirements, we move on to the actual selection, which takes place mainly based on the intensity and repetition of the channel-frequency pairs involved in the study: to avoid burdening the computational load in the online phases (as will be explained, we will have to calculate an high number of covariance matrices ) and mainly to reduce the over fitting, a reduced number of features is used, typically from three to maximum ten, focusing more on the stronger ones.

With all the features recorded, we move on to the creation of a feature vector (matrix size:

windows x number features) containing all the values and the corresponding indices that identify the task to be provided to the machine learning algorithm for the construction of the classifier: this one should be able to establish a set of parameters that maximise the statistical distance between the distributions of the two classes in order to minimise the levels of indecision over one assignment rather than another (*Figure 3.9*). From several studies [45] [46], the LDA in its linear variant (in case of covariance matrix of the features of the two equal classes) and quadratic (matrices of different covariance) effective especially for the quasi-gaussianity of the PSD of the two classes (hypothesis necessary for the application of the algorithm), are proved particularly useful in MI classification.




(a) Good distributions (b) Bad distributions

Figure 3.9: Figure *a* shows an example of well-separated distributions differently from *b*. A nice division leads to better classification from the model.

Of course, the performances, however good a single model is, are not enough to define its usefulness for a correct use of BMI and for this we resort to validation methods for the identification of the best possible parameters: when a classifier is created, the algorithm computes the statistical distribution of the two classes and plot them in a graph; moreover, the rejection in each calibration run is evaluated. This parameter shows a brief review of the way in which the classifier predict, with an accuracy lower than 55%, the sample outcome: this probability has a very low evidence (quasi-random classification) and, consequently, the model can not be used for a correct real-time prediction. Among all the classifier that can be created with the calibration run, the best is the one that shows the greater separation between classes and the less level of rejection. Other important metrics to define the degree of goodness of a classifier can be summarized and represented through the confusion matrix: inside it are listed the percentages of correct (respectively TP for the first class and TN for the second) and wrong (FP-FN) classi-

fications: these values can be used to understand the ability of the algorithm to differentiate the two tasks and to be able to make a correct classification; from these four parameters it is then possible to calculate the accuracy, or the overall capacity of the model to implement a correct classification.

Concerning this last point, it is necessary to pay particular attention to the phenomenon of the over-fitting: if the accuracy of the classifier is very high (over 90%, never achieved with the models created in this experiment), there is the risk that the model is highly dependent only on the data with which it has been created and, therefore, this would result in a very high miss-rating, resulting in errors in decoding the user's intentions in the online BMIs [47].

### 3.5.5 Accumulation framework

The output of the model when data are passed to the algorithm are samples, i.e., the classification takes place on the individual time samples of each window of the PSD: the outcome performed by the model is usually not continuous, generating task's prediction not suitable with the frequency of kinesthetic imagination. This event happens due the bias of the classifier caused by the small differences between classes distribution generated from poor informative features or by the presence of high amplitude artifacts that were not removed correctly; moreover, the prediction does not take into account its results in previous time instants. It is clear that this type of discontinuous classification is not suitable for an online BMI use. To solve these problems, it is therefore necessary a method that is able to accumulate evidence, or identify with an high probability the will of the subject to execute a task or the other based on the information a posteriori: this methodology is defined accumulation framework (AF), and it is generally based on one or more mathematical functions, dependent on parameters directly modulated by the researcher in order to optimize the classification performance as much as possible. Different types of AFs can be used to perform the online classification, but in this experiment an algorithm based on Exponential Smoothing function was used: specifically, the outcome varies its value in dependence on the chosen smoothing parameters and from the value of probability in output from the classifier in the current and previous times(*Figure 3.10*).

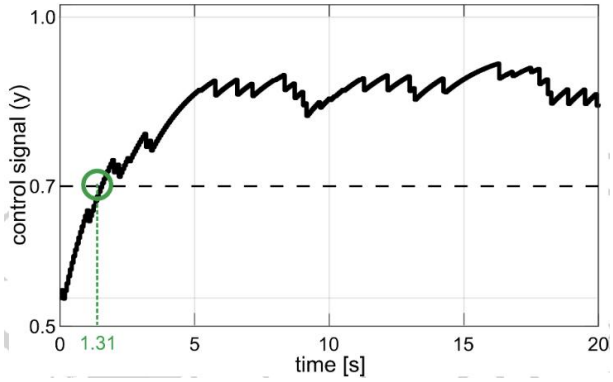



Figure 3.10: Example of the output of the exponential smoothing during time: from 1.31 s it is possible to notice the unstable behaviour of the prediction.

It is frequently important to find a balance to current prediction and those made in the distant past. Exponential smoothing, which uses previous calculated decision to create predictions, modulates its outcome according to value $\alpha$: this specific parameter (that can be chosen accordingly to the performance of the subject) modulates the "speed" by which the system accumulates evidence. When the output of the accumulation framework will have reached or exceeded the selected threshold (changeable by the scientist) to determine the class, the chances of the tasks will be reduced to 50%, the minimum value within which the percentage can not descend due to the duality of the classes, and the command will be sent to the BMI (for example, turn right or left depending on whether the progression completed is both hand or both-feet). The choice of such parameters can greatly influence the user experience in the control of BMI, facilitating the completion of a task compared to another (in case of difficulty) or instead making it more difficult, both to give the subject more control in the timing, and to reduce the effect of miss-classifications.

$$D(t) = D(t-1) * \alpha + pp(t) * (1-\alpha) \tag{3.3}$$

This last consideration is a sore point for BMI: in fact, if the user's neuro-physiological patterns are not stable during the experience, then the model fails to identify a correct discriminator between a class and another, producing a bias dependent on a priori parameters originated in its formulation, which leads to a considerable difficulty in completing one task and consequent shift to the other, albeit wrong. Moreover, the exponential smoothing can be particular unstable

when there are many miss-classified samples and it can not process correctly the absence of motor imagery. These fluctuations can be curbed both by choosing a more robust framework (e.g., [48]), with the progressive analysis of the features and with a subsequent re-calibration: if the model appears no longer able to follow the user's intentions, it must be re-calibrated, or correct it on the basis of the new information obtained, by keeping the old or recreating it from scratch.

Among the three days decided for the experiment, the first will then serve for the preliminary evaluation of the neural correlates of the subject and for the creation of a first classifier; during the second day we will proceed to test the model created, re-calibrating it if necessary with the features coming from the online test (update) or directly from new offline (creation of a new one); the third day will finally serve only as a test and the classifier will not suffer, except for cases that will be described in the next sections, no changes, even if its performance will be worse than previous uses.

## 3.6 Data analysis

Since neural correlates are one of the most important characteristics for the correct use of BMI, it is necessary to evaluate not only the performance of the classifier, related to the machine learning field, but mainly the parameters related to brain responses and its neuro-plasticity. Therefore, two mathematical methodologies linked to this neuro-physiological process have been considered, with the primary aim of understanding how the brain responds to the imagination of one task compared to another, with particular attention to how a different type of training, can affect the ability of the subject.

### 3.6.1 BMI performances

In the first instance, the training level of the subjects was evaluated directly with the numerical data obtained by the classifier and the performances of the individual days in the use of BMI. It has been analyzed the trend, run per run and day after day, of the sample accuracy (the percentage of correct classification of the single samples, dependent only on the classifier, SA), of the

command accuracy (percentage of tasks completed correctly, depending on both classifier and parameters chosen for the accumulation framework) and the required action completion time. These values were obtained for both types of training using the models and their evolutions during the three days. The SA has been analyzed for both offline and online, while the other two only for online.

### 3.6.2 Features stability

The usefulness of a ML approach can be quantified by analyzing the stability of the selected features to create or re-calibrate the classifier: the presence of one or more equal frequency-channel pairs for different runs indicates that the user prefers the modulation of those specific features and is therefore able to activate the brain areas in the same way despite the use of BMI for discontinuous periods of time. These values were not entrusted with a precise study in this paper, because the whole experiment is based on them and the assessments were made during the first two days of training as a metric to understand whether or not a subject would be able to control BMI. In fact, the subjects who managed to complete all three days were also those with stable features.

### 3.6.3 Fisher score and Riemann space

The ability to differentiate one class from another was mainly quantified through the study of the FS coefficient, described in detail in the previous sections, and by the Riemann distances[49]: this parameter studies the distances (in a new geometric space) between class using the covariance matrices. The EEG data of the channels involved in the MI were filtered into the useful bands and, similarly to what happened for the PSD, divided into the same number of windows according to the same mode described above; then, for each of these windows, a covariance matrix was created: in this way, through the Covariance Toolbox of MATLAB, it was possible to calculate the average distances for the classes in each run by first identifying the barycenters of the classes in the geometric space of Riemann (riemman mean) and the distance of these

values (distance riemann, RD) according to the formula

$$RD = \sqrt{\sum \log \lambda(a,b)^2} \tag{3.4}$$

where $\lambda$ represents the eigenvalues of the covariance matrix, *a* and *b* the two barycenters of the classes.

The FS was calculated with a double average between frequencies and channels involved in the study, in order to obtain a value per run of $\mu$ band and $\beta$ band.

To assess the evolution of the neuro-physiological patterns of the users, and to quantify the differences among them, we have not only based on the two indices just described, but also on their alternative versions that went to consider, respectively as a statistical difference and as a distance in the Riemann plane, the individual classes and their change during the sessions with respect to their very first value obtained for each subject: a total of six variables, two between classes, were analysed (classic versions of Riemann and Fisher) and four intra-class.

# Chapter 4

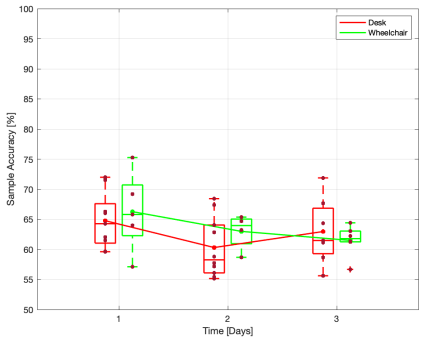
# Results

In this section we will discuss the results obtained through previous analyses: to get a better understanding about the main differences between one group and another, each consideration will be accompanied by non-parametric statistical Wilcoxon test (that analyzes the medians of distributions), to assess its reliability.

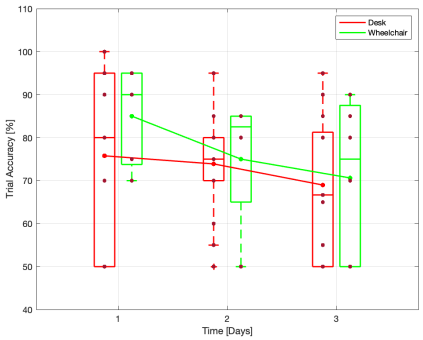
## 4.1 BCI-related metrics

The statistical metrics of trial accuracy, sample accuracy and delivering time are reported in the *Figure 4.1* together with the coefficients obtained by statistical tests for each of the days. The data of each subject were clustered together according to the group (on desk and on wheelchair): then, the statistical differences among them were evaluated according to the Wilcoxon test with a threshold of evidence set to $p_{value} < 0.05$. The median of wheelchair trial accuracy are slightly superior respect to the other one and also the means are different ($p_{value1} = 0.797$, $p_{value2} = 0.303$, $p_{value3} = 0.945$, no statistical difference found among the two groups): actually, these value see a difference, between the two training types, of $10.8\% \pm 19.5\%$ % during the first day, $1.48\% \pm 11.46\%$ during the second day and of $2.5\% \pm 15.74\%$ on the last day. For the sample accuracy, all the values that are less than 55% are not considered (classifier rejection): nevertheless, it is possible to see a difference between the averages of $2.2\% \pm 3.21\%$ in the first day, $4.3\% \pm 4.38\%$ in the second and $3\% \pm 5.2\%$ in the third but still no statistical evidence in any of the days. Finally, analyzing the average time of delivering commands, we see a statistical difference in favor of the on desk group on the first day ($p_{value} = 0.035$) but absent in the second and third day: unlike the other two variables, the members of the group in the wheelchair have higher values that undergo an improvement in the following days ($37\% \pm 3.41\% - 2.34\% \pm 1.77\% - 7\% \pm 7.80\%$). For all the metrics, it can be assess that the group on the wheelchair

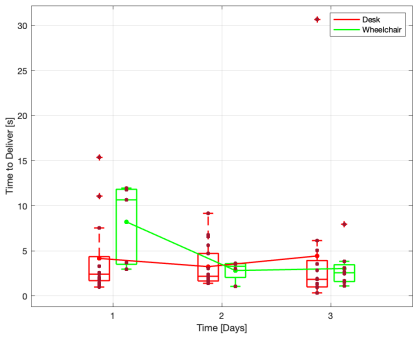
generally has fewer outliers than the other.




(a) Sample accuracy




(b) Trial accuracy




(c) Delivering time

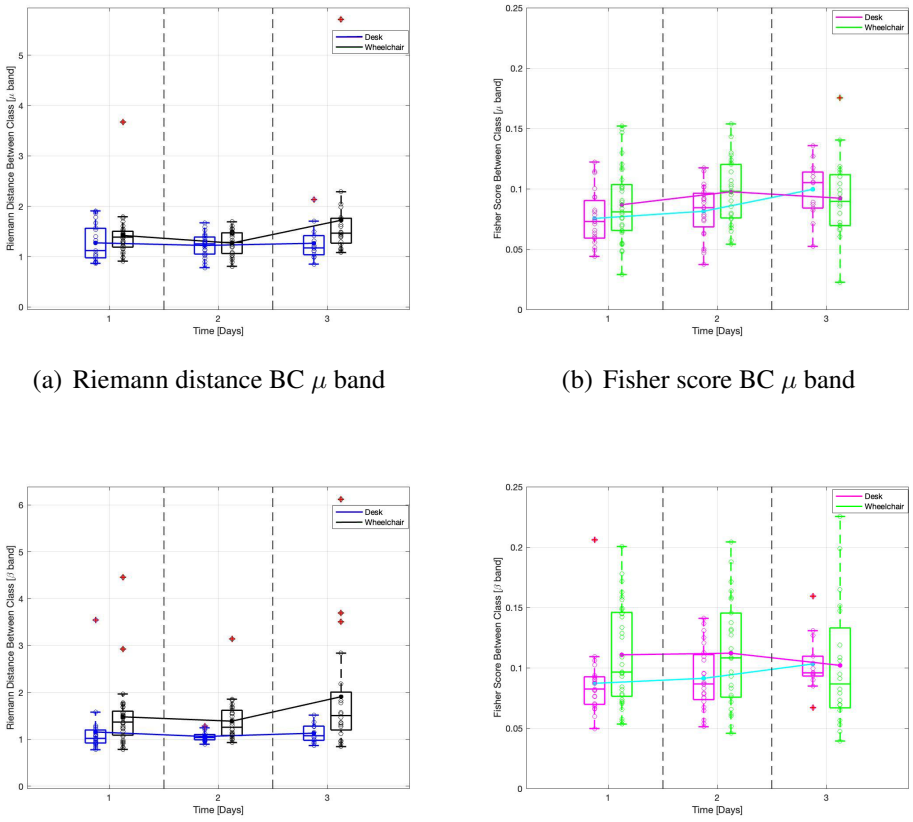
Figure 4.1: Boxplots representing data distribution of the two groups, during the three days, according to the metrics previously defined.

## 4.2 User learning

### 4.2.1 Between classes

The learning abilities of the subjects were evaluated mainly through the analysis of the FS and RD, both referring to the two MI classes. Since the typical frequencies of this mental task correspond to the $\mu$ and $\beta$ bands, all the considerations have been made about them, highlighting the main difference between the two, especially in terms of features used to calibrate the classifiers. As with the BCI metrics, also in this case the variations of the Fisher and Riemann coefficients were evaluated as the days varied for both frequencies and study groups.

The Riemann distance in the $\mu$-band (*Figure 4.2*) evolved over the three days, showing that the distributions were statistically equal between the two groups in the first two days ($p_{value1} = 0.1521$, $p_{value2}$ = 0.4765) but were different in the third day ($p_{value3} = 0.02$) with higher values for those who started training directly in the wheelchair. On the other hand, the evolution of the coefficients in the beta band was quite different, since on each day the distributions were statistically different ($p_{value1} = 0.0021$, $p_{value2} = 0.0011$, $p_{value3} = 0.0136$) with very high evidence and significantly higher values in the wheelchair group. Analyzing the FS in the mu band, it is possible to notice a slight increase in the discrimination between the first and the second day for both classes (about 5%) with a pairing of the group on wheelchair compared to the group on desk on the third day (difference between them of 0.8%): under a statistical analysis, only the second day showed evidence of difference, with a $p_{value} = 0.0351$. In the beta band, instead, there is a decrease in the values of the second group and an increase in those of the first, with only the first day with an high statistical difference ($p_{value} = 0.0344$), although visually it is still possible to notice higher average values in the first two days and equal in the third.




(a) Riemann distance BC $\mu$ band

(b) Fisher score BC $\mu$ band

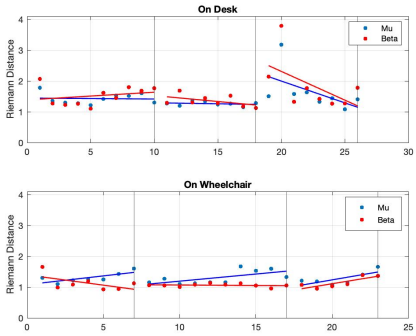
(c) Riemann distance BC $\beta$ band

(d) Fisher score BC $\beta$ band

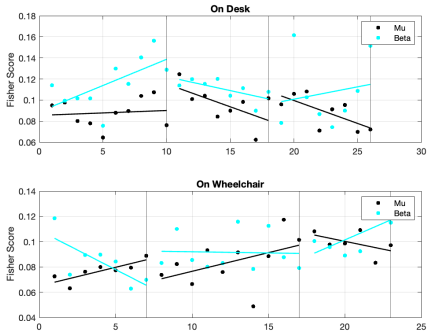
Figure 4.2: Boxplots representing data distribution of the two groups, during the three days, according to the differences between classes.

To study more accurately the evolution of these indices, it is advisable to focus not only on the three days, but also on the individual runs, so as to understand which regression model learning it follows and whether the presence of outlier data can lead to a shift in statistical considerations. To do this you can rely on the population analysis method Naive Average: its use is recommended only when the data available are of the same number for all subjects (here the mean is biased towards subjects, like a8, that had more runs for days respect to the others) and, for this reason, it is used just to have a representation of the general evolution. The trend of the linear regression (*Figure 4.3*) is very close to the real data, showing an averaged difference wheelchair-desk of $9.3\% \pm 0.32\% - 3.4\% \pm 0.70\% - 28.6\% \pm 0.65\%$ in $\mu$ band and of $34.9\% \pm 0.14\% - 27.6\% \pm 0.23\% - 60.1\% \pm 1.12\%$ in $\beta$ band (Riemman). Moreover, the outlier are almost totally disappeared for wheelchair dataset while for desk there are still

four that make interpolation more optimistic. As you can see, the regression of the group on wheelchair follows very faithfully the data, especially on the third day, differently from what happens to the other group whose data have a certain variance between them, much more drastically remarked during the last training session. In addition, you can see that the data of the second group are highly correlated to single runs for the second day in $\mu$ band ($p_{value} = 0.04$ - $\rho$ = 0.65) and for $\beta$ band in the third day ($p_{value} = 0.04$ - $\rho$ = 0.83), which is not the case for the first group. Different considerations appear for the same analyses performed on the FS: the standard group does not show any high statistical evidence correlation on any of the days except for first day in $\beta$ ($p_{value} = 0.04$ - $\rho$ = 0.64), while the other group only have it on the first day in $\mu$ ($p_{value} = 0.03$ - $\rho$ = 0.82); moreover, the data have no linear behavior (if not for the first day of wheelchair); in general you can see an average difference between groups of $12.7\% \pm 0.67\% - 12.3\% \pm 0.78\% - 11.6\% \pm 0.66\%$ in the $\mu$ band and $27.7\% \pm 0.69\% - 17.0\% \pm 0.55\% - 2.1\% \pm 0.39\%$ in the $\beta$ band.




(a) Averaged Riemann distance BC



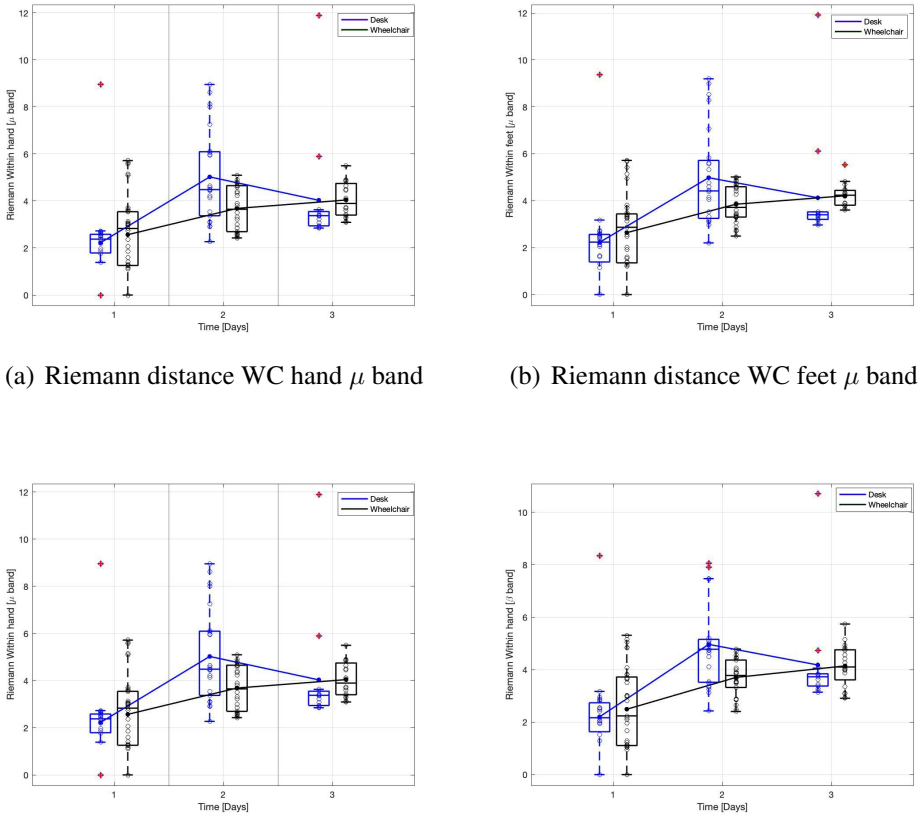

(b) Averaged Fisher score BC

Figure 4.3: From this graphs the information on the evolution of indices can be noted. The total number of point is different between the two groups due to different number of runs acquired during the days.

### 4.2.2 Within classes

Another analysis that can be carried out is based on the evolution of the coefficients referred to the individual classes taken separately: for a correct vision and evaluation, the same techniques for the analysis between classes will be used . The evolution of the classes according to the Riemman metric, is quite different between the two groups in the two frequencies, as

the distances of the group on wheelchair tend to grow considerably on the second day and then decrease in the third, while those of the control group have a much slower but continuous increase, resulting, compared on the final day, average equal in all cases except for the class feet in the beta band (difference of 11.5%). By the application of the Wilcoxon test, it appears that for the hand class, there is a significant statistical difference between the two groups for the second day in both bands ($p_{value_{mu}}$ = 0.027, $p_{value_{beta}}$ = 0.002) and for the third only in the mu band ($p_{value}$ = 0.026), while for the feet class, the third day is only significant for the mu band ($p_{value} << 0.05$) and the second only for the beta ($p_{value} << 0.05$). In addition, both groups have a large number of outliers, especially in the beta band (*Figure 4.4*).




(a) Riemann distance WC hand $\mu$ band

(b) Riemann distance WC feet $\mu$ band

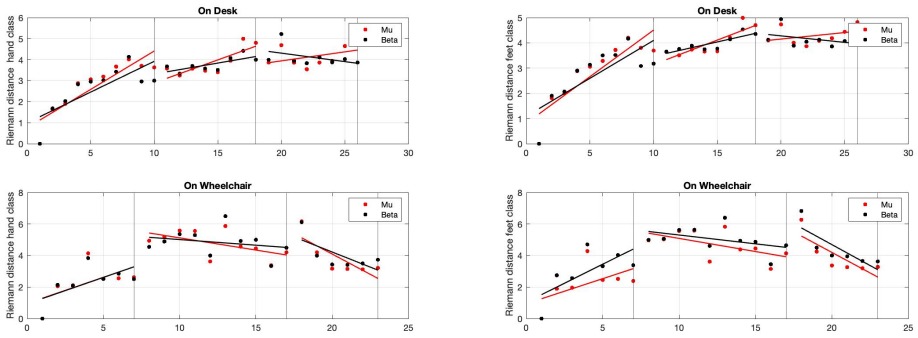
(c) Riemann distance WC hand $\beta$ band

(d) Riemann distance WC feet $\beta$ band

Figure 4.4: Boxplots representing data distribution of the two groups, during the three days, according to the differences between classes.

Analyzing the average trend with respect to the runs, we note a continuous growth of the

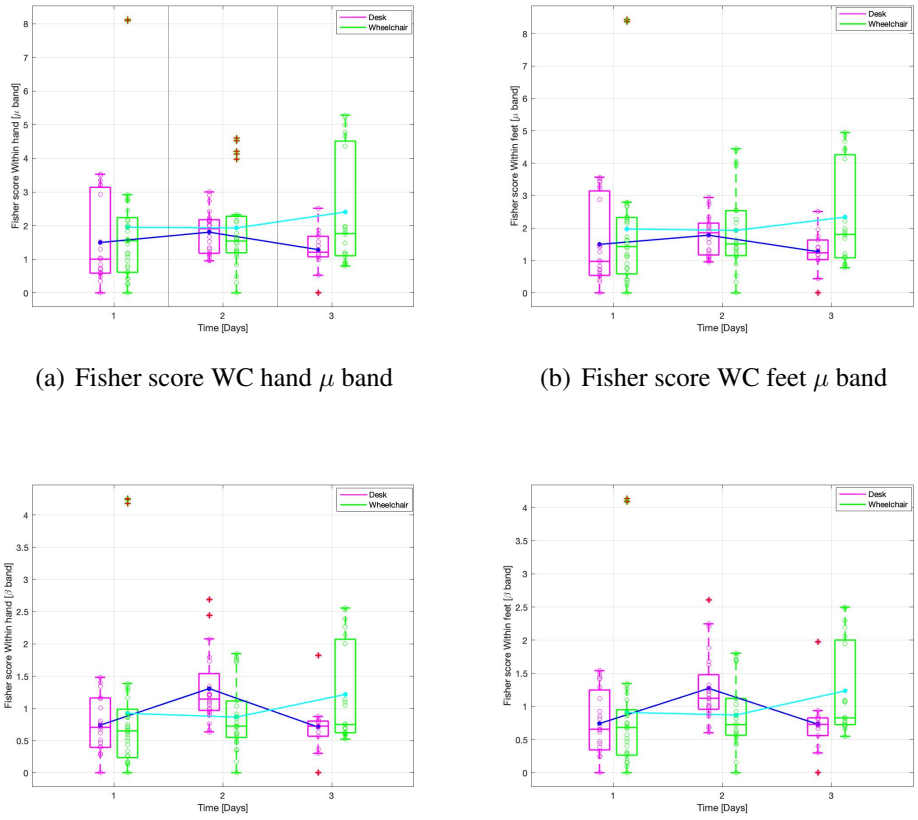
coefficients of the control group and instead a decrease, for the two bands, regarding the third day of the second group; for the desk group, the data are highly correlated to the runs in $\mu$ ($p_{value1} << 0.05$ - $\rho_1$ = 0.88, $p_{value2} < 0.05$ - $\rho_2$ = 0.8) and in $\beta$ ($p_{value1} << 0.05$ - $\rho_1$ = 0.78, $p_{value2} = 0.04$ - $\rho_2$ = 0.72) for the first two days for hand class; in the feet class, there are the same correlation with same values. The second group instead have only a third day correlation ($p_{value}$ = 0.047 - $\rho$ = -0.82) in the beta band for class feet. In addition, the data appear very disordered and the regression line is therefore not very informative on the correct trend, assumption that is also deduced from the correlation values. The evolution of these coefficients seems to follow a crescent-plateau trend in the first two days (*Figure 4.5*).




(a) Averaged Riemann distance WC hand (b) Averaged Riemann distance WC feet

Figure 4.5: From this graphs the information on the evolution of indices can be noted. The total number of point is different between the two groups due to different number of runs acquired during the days.

In relation to the FS, it is possible to note that for the $\mu$ band the FS within hand do not have any statistical difference between groups (averaged $p_{value} = 0.45 \pm 0.27$) while, for the $\beta$ band, there is a difference only during the second day ($p_{value} << 0.05$). Instead, according to the feet class, the two groups differ statistically only for $\beta$ band in the second e third day ($p_{value2} << 0.05$, $p_{value2}$ = 0.037) (*Figure 4.6*).




(a) Fisher score WC hand $\mu$ band

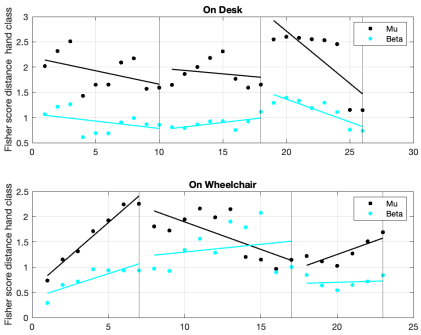
(b) Fisher score WC feet $\mu$ band

(c) Fisher score WC hand $\beta$ band

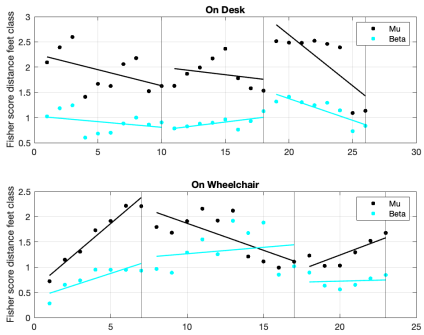
(d) Fisher score WC feet $\beta$ band

Figure 4.6: Boxplots representing data distribution of the two groups, during the three days, according to the differences within classes.

Finally, the averaged evolution (*Figure 4.7*) seems much worse than Riemman: the data, for desk group and both bands, at the start of every day follow a semi-linear trend but towards the end of the day they decrease their value, especially for the second and third days. For the hand of desk group, the data on $\mu$ band are correlated with the runs on the third day ($p_{value} = 0.02$ - $\rho$ = -0.78) and in second and third day for $\beta$ ($p_{value2} = 0.08$ - $\rho_2$ = 0.64, $p_{value3} =<< 0.05$ - $\rho_3$ = - 0.87); for feet class happens the same behaviour with a correlation in $\mu$ band on the third day ($p_{value} = 0.02$ - $\rho$ = -0.78) and on beta for second and third day ($p_{value2} = 0.07$ - $\rho_2$ = 0.66, $p_{value3} << 0.05$ - $\rho_3$ = - 0.85). Regarding the second group, class hands is highly correlated in $\mu$ band for the last two days ($p_{value2} = 0.02$ - $\rho_2$ = -0.72, $p_{value3} = 0.05$ - $\rho_3$ = - 0.81) but only in the first day for $\beta$ band ($p_{value} = 0.02$ - $\rho$ = 0.85) .




(a) Averaged Fisher score WC hand



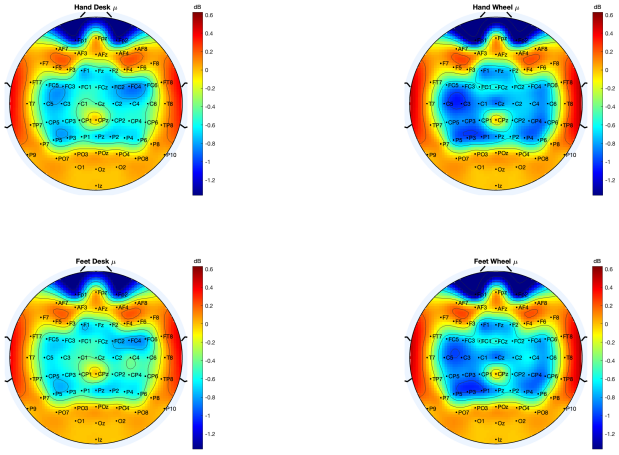

(b) Averaged Fisher score WC feet

Figure 4.7: Data follow a general linear trend but with some problems in desk group, visible on second and third days.

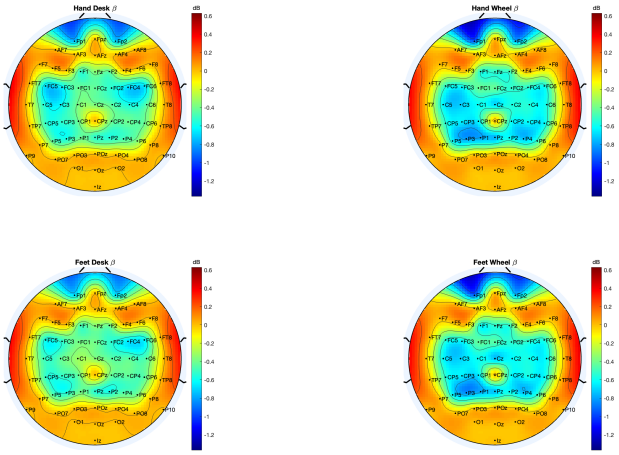
### 4.2.3 ERD/ERS topoplots

From a topographical analysis of the average activity of the subjects in the three days, it is possible to have a general view of the differences between the group on desk and the group on wheelchair: in particular, it is interesting to observe how the electrodes that register the desynchronization (indicating the neural activity), vary depending on the task (both hand, both feet) considered in the two bands.

From the *Figure 4.8*, it can be assess that the ERD in $\mu$, for the wheelchair group, is generally more marked, especially in the electrodes C5-C3 for both taskes but, at the same time, also Cz is highlighted in hand task for this group, even if it refers to feet movement; moreover, also the parietal areas show an high desinchronization. For $\beta$ band, the map shows higher desinchronization for wheelchair group, especially on the parietal areas; moreover, in this case, Cz is present only in feet task and only in the wheelchair group.



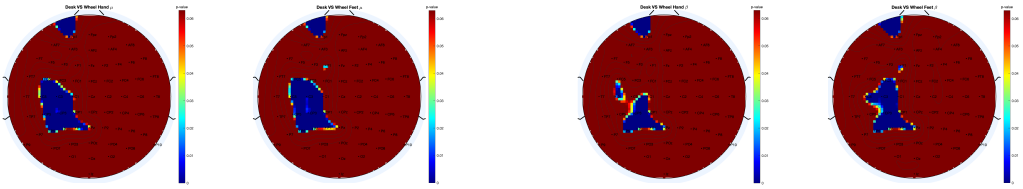

(a) $\mu$ band topoplot




(b) $\beta$ band topoplot

Figure 4.8: ERD/ERS maps for the two groups and the two task. The data are expressed in dB and they are the results of interpolation from 32 to 64 electrodes to enhance the quality. The blue color means ERD, the red ERS.

The statistical differences, according to t-test, are shown in *Figure 4.9*: for both tasks, in $\mu$ band, the electrodes C5-CP5-C3-CP3 are statistical different with a $p_{value} = 0.01$ while, for feet class, it is possible to see a difference also in the F1 electrodes ($p_{value} = 0.03$). Instead, in the $\beta$ band, the statistical different area are that of electrodes CP3-C5 for hand task ($p_{value} = 0.01$) and of FC3-C3-CP3-C5-P5 for feet task ($p_{value} = 0.01$); even in this case, for feet task it is possible to see a difference with $p_{value} = 0.02$ in the F1 electrode.



(a) $p_{value}$ map for $\mu$ band

(b) $p_{value}$ map for $\beta$ band

Figure 4.9: p-map of the active electrodes. The statistical difference is set when the $p_{value}$ is lower than 0.05 and it is represented with the blue color.

### 4.2.4 Spectrograms

As explained above, most subjects modulate only certain sub-bands in the MI and, for this reason, it is recommended to use the PSD as a discriminating feature. In addition to the topographic maps, the average spectrograms of the two groups were analysed during the two MI tasks: on average, the task "both hand" lasted $1.7 \pm 1.2s$ for the group on desk and $2.1 \pm 0.85s$ for the group on wheelchair, while the task "both feet" $1.48 \pm 0.56s$ against $1.88 \pm 0.63s$. Such time windows are different from each other, and therefore it was chosen to analyze the mean spectrograms in the time intervals of 1.9s for the class "both hands" and 1.7s for the class "both feet"; moreover, the graphs are those obtained by averaging the information from the same channels used in the RD e FS. In the *Figure 4.10*, the spectrograms of the two groups for the two tasks can be shown: it is possible to notice that generally, for the wheelchair group, the averaged PSD is lower in the $\beta$ band, according to the features used for the classifiers, and it is discontinuous with gaps of 0.2s in both tasks.

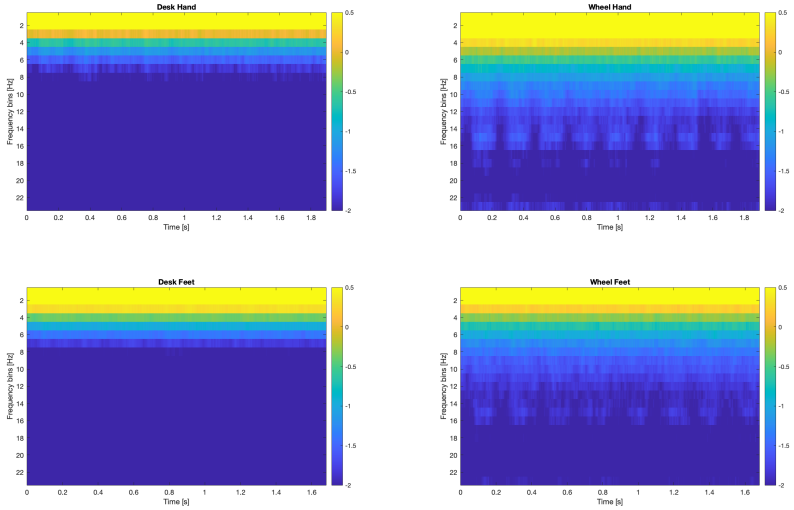



Figure 4.10: Averaged PSDs in the selected channels. The frequency bins have a resolution of 2 Hz and this value match with the PSD created for the classification.

Remarkably, even if the dissimilarities among the maps are not completely notables, the application of t-test shows extremely high differences for most of the frequency-time pairs (*Figure 4.11*): the almost totality of the map of "both hand" shows statistical difference between the two groups, with an average $p_{value} = 0.003 \pm 0.007$; also the task "both feet" shows the rejection of null hypothesis, especially in the range 6-22 Hz (average $p_{value} = 0.002 \pm 0.006$),

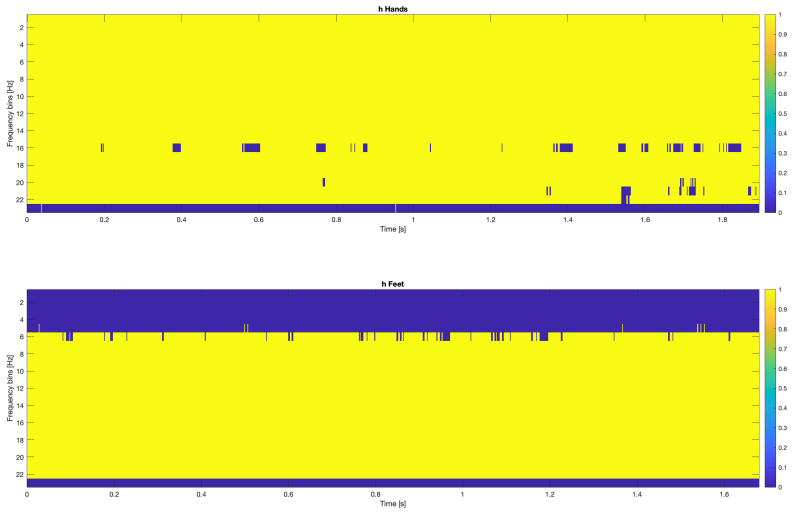



Figure 4.11: Hypothesis rejection map according to t-test: the yellow pairs shows its rejection

### 4.2.5 Features maps

From the analysis of the selected features for the creation of the classifier, it is possible to understand the differences between the most discriminant channel-frequency pairs for the subject: those maps, showed in *Figure 4.12*, are the results of the average of calibration run used for the creation of the classifier. The more discriminant features appear on the correct frequency-channel pairs on the desk group while, for the wheelchair group, there are features in the right channels but for extremely high frequencies.

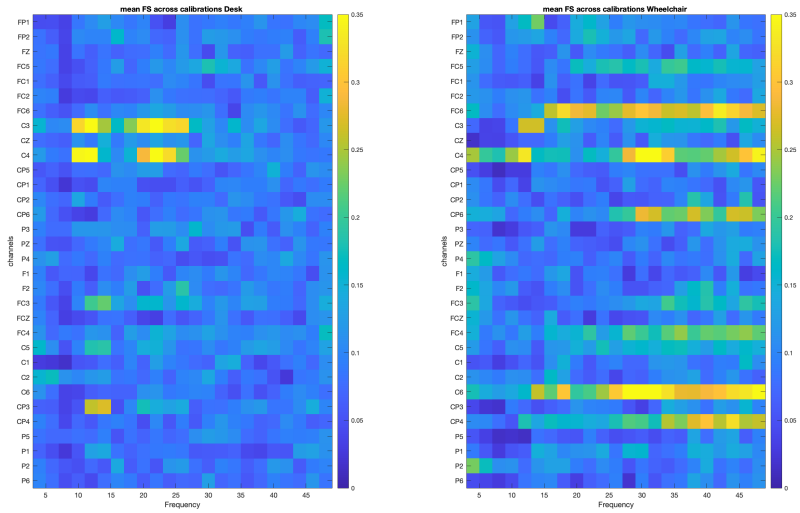



Figure 4.12: Averaged Fisher Maps of calibration files for group on desk and on wheelchair.

Moreover, a t-test can be applied to verify the statistical differences between the two groups (*Figure 4.13*): actually, the h for the most part of MI channels and frequency show some differences ($p_{value} = 0.024 \pm 0.01$), especially in the $\mu$ band.

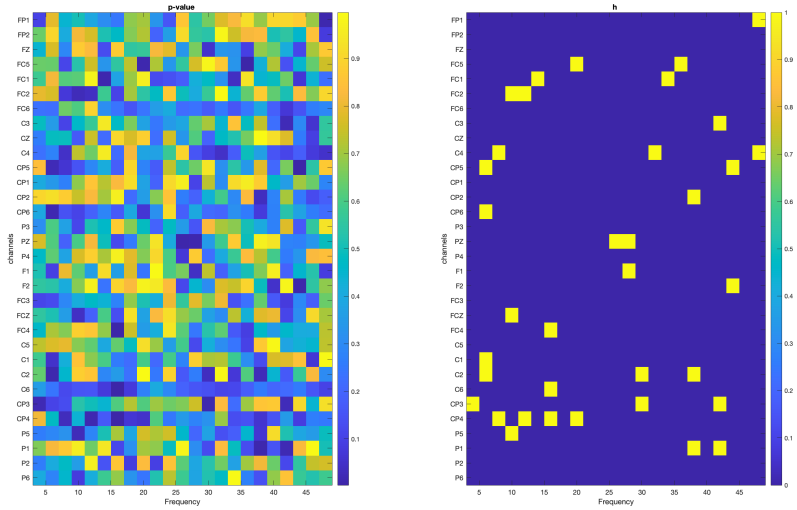



Figure 4.13: Hypothesis rejection map according to t-test: the yellow pairs shows its rejection

# Chapter 5

# Discussion

In this study we focused, for the first time, on analyzing the effect that a different training than the standard one of BMI can have on the ability of an individual to control an external device. From the results listed above, a good difference appears between the two groups not so much in the performance of the algorithm itself, as in the neural correlates of the individual subjects: in fact, the RD BC showed how the members of the group on wheelchair were able, already from the second day, to differentiate themselves positively from the other group in the distinction of the two classes, mostly in the $\beta$ band. In this regard, during the study, it was noted that 3/4 of the group on desk showed features in the first $\beta$ bands while those of the second group used only the $\mu$ bands, indicative situation of greater control over brain modulation. On the other hand, however, the FS does not lead to the same considerations: from the plots in the runs,it can be assess that the data are quite unstable and with an high scatter for both groups, especially for the presence of outliers that, even with the use of averaging approach, do not completely disappear. These trends are in line with the variation of the accuracy of the subjects: in fact, the performances of the classifier are quite fluctuating, showing very low values (50%) even if the subject has shown excellent skills in previous runs; moreover, despite the channel-frequency pairs remained the same, sometimes it was necessary to re-calibrate the classifier on the same features because the intensity of brain modulation varied day after day.

An excellent functional analysis can be made mainly for the evolution of within indices for both Fisher and Riemann: although they are not directly considered to understand how to improve BMI and their algorithms, this information is very important to understand how brain activity changes with training and, above all, how the user approaches the same task over time. Even with a simple visual analysis it is possible to notice that the two classes have similar evolu-

tions (meaning that, considering Riemman Space, they move parallel and with always constant distance between them) both in the distributions, as can be deduced from the boxplots, and during the runs. Moreover, the second group on the last day tends to return to values similar to the initial ones. This factor actually has an important bearing on ML theory since it states what was proposed in the previous sections: after all, one of the goals was to assess whether it was possible to focus on user patterns and not only on the algorithm itself. A return to initial values could indicate a significant increase in the ability of the individual to better modulate brain activity and, above all, an effort to maintain these features constant over time, also for a long period beyond that proposed for the experiment; moreover it can be found that the differences, even if minimal in all the variables, favor the training directly with the final goal (whether it's the wheelchair as in this experiment or another machine as a robotic arm) as the results show greater control of BMI, especially in the absence of the first beta waves.

Finally, from the topoplots, an important evaluation of the ability of learning of the subjects can be evaluated. First of all, it can be seen that the group on wheelchair shows high ERD values in the P electrodes, while they are not present in the group on desk: actually, the parietal lobe has the function to process the information of outside movement ([24]) and, due to the rotation of wheelchair, this area can be activated during the last period of each correct task (desinchronization are more evident for the $\mu$ band, and this is consistent with the preferred band for the wheelchair group). An other important characteristic is the presence of ERD in the Cz electrode, only for the group on wheelchair, meaning that its subjects are able to modulate their activity also for "both feet" task that is generally difficult for most of the users; additionally, the statistically different areas are those related to the motor cortex on the left hemisphere (all the subjects were right-handed).

In conclusion, we must also take a look at the characteristics and performance of individual subjects mainly in the third day: among the 8 subjects, only 5 have completed the training effectively, 3 of them (a8, a5, c1) belonged to the first group and the remaining 2 (c5, c4) to the second group, although this number is in line with the typical percentages of BMI use,

it is appropriate to consider the bias of the results (also because in many days, there was not enough silence to allow subjects to concentrate sufficiently, a sign visible even from the high frequencies highlighted by Fisher's maps). This bias is extremely evident on the Fisher maps because the averaging of the wheelchair group considered only the two subjects that made just one calibration, and for this reason the values are not cleaned enough. All users who did not arrive at the end of the experiment encountered difficulties from day one in the control, many of them stating that they could not fully have the feeling of "presence" arts or find it difficult to imagine the same task continued for longer. This limit, however, is not fully detectable (see individual graphs in the appendix) and needs some clarification in order to have a better understanding of how the closed loop user-computer-machine works. First, the data of the FS are obtained from the average of the values in each run of frequencies/sub-bands and channels chosen and therefore, the presence of features with very high values overestimates the final index: the value considered in the graphs in the results, not only takes into account the features used for calibration but all those that, from literature, could be involved in the MI and this implies that, even if a subject has the ability to use the BMI, his features are not constant over time, making it impossible to build a classifier but still resulting in good values for general analysis. Additionaly, for a8 and c1 it was necessary a calibration also on the third day because their performances were worst respect to the other days. Finally, we must also take into account the confidence levels that have been taken despite the first approach to BMI: starting directly in the wheelchair, the members of the second group were able to learn more about how it worked and thus minimise the surprise effect of rotations.

# Chapter 6

# Conclusion and outlooks

The main objective of this paper was to show how a training based on ML and with a direct approach to the device to be controlled, can bring countless benefits in the use of BMI and in its role outside the research field. The results shown and discussed so far confirm that a training based on ML and organized with the approach, from the first day, with the wheelchair, improve the ability of the users to use BMIs compared to the ones that trained in the standard way. The main contribution of my thesis work was to provide further insight into the inter-individual characteristics that control brain modulations, highlighting how users are able to adapt gradually to what is proposed, thus making faster and more efficient decoding of their intentions, even after a prolonged period of time. This ability to keep the features constant is extremely important, especially in order to minimize possible temporary fluctuations that, otherwise, would not allow a proper usage; identifying the suitable areas is therefore also possible to increase the range of individuals (also affected by diseases) that can access BMI, thanks to a clear reduction in the time needed for training and, consequently, mental fatigue.

It was then shown that many of the differences between the groups resided in RD (reviewed in literature [50] to be a method for channel selection), although this parameter was not taken as a possible feature for the classifier: this suggests that a further step forward could be to combine the PSD-based methodologies used in this experiment, with the information obtained from temporal covariance matrices, identifying algorithms that can reduce the online computational weight for such calculations. The possibility to exploit features coming from different spaces could allow greater flexibility in the creation of the classifier and, above all, would allow to include the information obtained from this study on the effect of ML in the neuro-physiological evolution.

Most importantly, although not all subjects were able to complete the experiment, they all showed comparable interesting values: this is a further sign that you could increase the pool of users able to use BMI simply by incorporating new information, both in the creation of models, both in the ways of teaching them to maximize the experience. However, some subjects showed better performances than others, and in particular two of them, b1 and c4. Both showed noteworthy peculiarities: while the first, belonging to the first group, was able to complete almost perfectly every run (with some drops) only through the very first created classifier, the second, belonging to the second group, has gone through more re-calibrations of the same features but he was able to complete the runs faster than all the others, still achieving excellent performances. It is clear that these results, despite being the best of the groups, confirm the importance of the stability of the features, neverthless without reducing the control capabilities.

Of course, for the purpose of practical use, the three days of training proposed for the experiment may not be enough for everyone and therefore a long-term study would be necessary, taking into account also the differences that could exist further diversifying the devices used, in order to see if the performances also depend on the degree of appreciation and interest that the user has in it. In addition, it would be appropriate to have a much larger number of subjects, in order to be able to assess the large-scale impact of these methodologies and to evaluate in a better way the population study, to identify common patterns and have more data available on which make more comparisons and observe if the success rates of the control tend to remain constant compared to this study or it may vary: regarding this, it should be noted that, despite the excellent results listed above, only two (40%) members of the wheelchair group had access to the final phase, compared to the three (60%) of the on desk group.

Finally, it would be of great interest to observe how the results of training for people with disabilities could vary from those proposed in this paper and introduce, in case of optimal results, such technology in the various levels of rehabilitation, depending on the needs of each individual subject [51].

# REFERENCES


[1] Gerard J Tortora. *Principles of human anatomy*. HarperCollins College, 1995.

[2] Bruce P Bean. The action potential in mammalian central neurons. *Nature Reviews Neuroscience*, 8(6):451–465, 2007.

[3] Sanford L Palay and George E Palade. The fine structure of neurons. *The Journal of biophysical and biochemical cytology*, 1(1):69, 1955.

[4] Gregory L. Holmes and Roustem Khazipov. Basic neurophysiology and the cortical basis of eeg. *The Clinical Neurophysiology Primer*, pages 19–22, 2007.

[5] H. H. Jasper. The ten–twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, 10:367–380, 1958.

[6] Antoine Lutz, Lawrence L Greischar, Nancy B Rawlings, Matthieu Ricard, and Richard J Davidson. Long-term meditators self-induce high-amplitude gamma synchrony during mental practice. *Proceedings of the national Academy of Sciences*, 101:16369–16373, 2004.

[7] Matti Hamalainen, Riitta Hari, Risto J Ilmoniemi, Jukka Knuutila, and Olli V Lounasmaa. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of modern Physics*, 65:413, 1993.

[8] Martine Gavaret, Louis Maillard, and Julien Jung. High-resolution eeg (hr-eeg) and magnetoencephalography (meg). *Neurophysiologie Clinique/Clinical Neurophysiology*, 45:105–111, 2015.

[9] Seiji Ogawa, Tso-Ming Lee, Alan R Kay, and David W Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *proceedings of the National Academy of Sciences*, 87:9868–9872, 1990.

[10] Dale L Bailey, Michael N Maisey, David W Townsend, and Peter E Valk. *Positron emission tomography*, volume 2. Springer, 2005.

[11] Nathan E Crone, Alon Sinai, and Anna Korzeniewska. High-frequency gamma oscillations and human brain mapping with electrocorticography. *Progress in brain research*, 159:275–295, 2006.

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

[13] Noman Naseer and Keum-Shik Hong. fnirs-based brain-computer interfaces: a review. *Frontiers in human neuroscience*, 9:3, 2015.

[14] Niels Birbaumer. Breaking the silence: brain–computer interfaces (bci) for communication and motor control. *Psychophysiology*, 43(6):517–532, 2006.

[15] Jonathan R Wolpaw. Brain-computer interfaces (bcis) for communication and control. In *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, pages 1–2, 2007.

[16] J Vernon Odom, Michael Bach, Colin Barber, Mitchell Brigell, Michael F Marmor, Alma Patrizia Tormene, Graham E Holder, et al. Visual evoked potentials standard (2004). *Documenta ophthalmologica*, 108(2):115–123, 2004.

[17] Terry W Picton, Steven A Hillyard, Howard I Krausz, and Robert Galambos. Human auditory evoked potentials. i: Evaluation of components. *Electroencephalography and clinical neurophysiology*, 36:179–190, 1974.

[18] Denis R Giblin. Somatosensory evoked potentials in healthy subjects and in patients with lesions of the nervous system. *Annals of the New York Academy of Sciences*, 1964.

[19] Emanuel Donchin and Michael GH Coles. Is the p300 component a manifestation of context updating? *Behavioral and brain sciences*, 11(3):357–374, 1988.

[20] Martin Spüler, Wolfgang Rosenstiel, and Martin Bogdan. Online adaptation of a c-vep brain-computer interface (bci) based on error-related potentials and unsupervised learning. *PloS one*, 7(12):e51077, 2012.

[21] Dean J Krusienski, Eric W Sellers, Dennis J McFarland, Theresa M Vaughan, and Jonathan R Wolpaw. Toward enhanced p300 speller performance. *Journal of neuroscience methods*, 167(1):15–21, 2008.

[22] Jean Decety. The neurophysiological basis of motor imagery. *Behavioural brain research*, 77(1-2):45–52, 1996.

[23] Robert M Hardwick, Svenja Caspers, Simon B Eickhoff, and Stephan P Swinnen. Neural correlates of motor imagery, action observation, and movement execution: a comparison across quantitative meta-analyses. *BioRxiv*, page 198432, 2017.

[24] Sébastien Hétu, Mathieu Grégoire, Arnaud Saimpont, Michel-Pierre Coll, Fanny Eugène, Pierre-Emmanuel Michon, and Philip L Jackson. The neural network of motor imagery: an ale meta-analysis. *Neuroscience & Biobehavioral Reviews*, 37(5):930–949, 2013.

[25] Werner Poewe, Klaus Seppi, Caroline M Tanner, Glenda M Halliday, Patrik Brundin, Jens Volkmann, Anette-Eleonore Schrag, and Anthony E Lang. Parkinson disease. *Nature reviews Disease primers*, 3(1):1–21, 2017.

[26] Luz Maria Alonso-Valerdi, Ricardo Antonio Salido-Ruiz, and Ricardo A Ramirez-Mendoza. Motor imagery based brain–computer interfaces: An emerging technology to rehabilitate motor deficits. *Neuropsychologia*, 79:354–363, 2015.

[27] Nikolaus Weiskopf, Klaus Mathiak, Simon W Bock, Frank Scharnowski, Ralf Veit, Wolfgang Grodd, Rainer Goebel, and Niels Birbaumer. Principles of a brain-computer interface (bci) based on real-time functional magnetic resonance imaging (fmri). *IEEE transactions on biomedical engineering*, 51(6):966–970, 2004.

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

[29] Yi-Qian Hu, Tian-Hao Gao, Jie Li, Jia-Chao Tao, Yu-Long Bai, and Rong-Rong Lu. Motor imagery-based brain-computer interface combined with multimodal feedback to promote upper limb motor function after stroke: A preliminary study. *Evidence-Based Complementary and Alternative Medicine*, 2021, 2021.

[30] Elon Musk et al. An integrated brain-machine interface platform with thousands of channels. *Journal of medical Internet research*, 21(10):e16194, 2019.

[31] Amy L Orsborn, Siddharth Dangi, Helene G Moorman, and Jose M Carmena. Closed-loop decoder adaptation on intermediate time-scales facilitates rapid bmi performance improvements independent of decoder initialization conditions. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 20(4):468–477, 2012.

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

[33] Luca Tonin, Gloria Beraldo, Stefano Tortora, and Emanuele Menegatti. Ros-neuro: An open-source platform for neurorobotics. *Frontiers in Neurorobotics*, page 91, 2022.

[34] Gloria Beraldo, Stefano Tortora, Emanuele Menegatti, and Luca Tonin. Ros-neuro: implementation of a closed-loop bmi based on motor imagery. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2031–2037. IEEE, 2020.

[35] Claudio Carvalhaes and J Acacio De Barros. The surface laplacian technique in eeg: Theory and methods. *International Journal of Psychophysiology*, 97(3):174–188, 2015.

[36] Christa Neuper, Michael Wörtz, and Gert Pfurtscheller. Erd/ers patterns reflecting sensorimotor activation and deactivation. *Progress in brain research*, 159:211–222, 2006.

[37] Li Zhao and Yang He. Power spectrum estimation of the welch method based on imagery eeg. In *Applied Mechanics and Materials*, volume 278, pages 1260–1264. Trans Tech Publ, 2013.

[38] Maurice S Bartlett. Smoothing periodograms from time-series with continuous spectra. *Nature*, 161(4096):686–687, 1948.

[39] Jean Decety. The neurophysiological basis of motor imagery. *Behavioural brain research*, 77(1-2):45–52, 1996.

[40] Gert Pfurtscheller, Clemens Brunner, Alois Schlögl, and FH Lopes Da Silva. Mu rhythm (de) synchronization and eeg single-trial classification of different motor imagery tasks. *NeuroImage*, 31(1):153–159, 2006.

[41] Julie Carrier, Stephanie Land, Daniel J Buysse, David J Kupfer, and Timothy H Monk. The effects of age and gender on sleep eeg power spectral density in the middle years of life (ages 20–60 years old). *Psychophysiology*, 38(2):232–242, 2001.

[42] Quanquan Gu, Zhenhui Li, and Jiawei Han. Generalized fisher score for feature selection. *arXiv preprint arXiv:1202.3725*, 2012.

[43] Niko A Busch and Rufin VanRullen. Spontaneous eeg oscillations reveal periodic sampling of visual attention. *Proceedings of the National Academy of Sciences*, 107(37):16048–16053, 2010.

[44] Suresh D Muthukumaraswamy. High-frequency brain activity and muscle artifacts in meg/eeg: a review and recommendations. *Frontiers in human neuroscience*, 7:138, 2013.

[45] Jaime F Delgado Saa and Miguel Sotaquirá Gutierrez. Eeg signal classification using power spectral features and linear discriminant analysis: a brain computer interface application. In *Eighth Latin American and Caribbean Conference for Engineering and Technology*, pages 1–7. LACCEI Arequipa, 2010.

[46] Carmen Vidaurre, A Schlogl, Rafael Cabeza, Reinhold Scherer, and Gert Pfurtscheller. Study of on-line adaptive discriminant analysis for eeg-based brain computer interfaces. *IEEE transactions on biomedical engineering*, 54(3):550–556, 2007.

[47] Douglas M Hawkins. The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1):1–12, 2004.

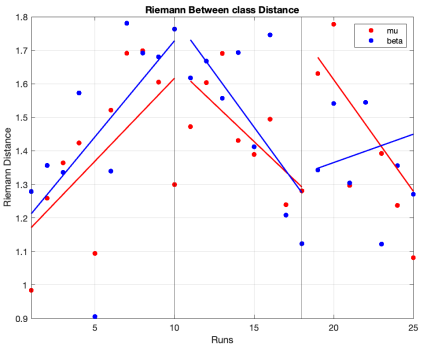
[48] Luca Tonin, Felix Christian Bauer, and José del R Millán. The role of the control framework for continuous teleoperation of a brain–machine interface-driven mobile robot. *IEEE Transactions on Robotics*, 36(1):78–91, 2019.

[49] Marco Congedo, Alexandre Barachant, and Rajendra Bhatia. Riemannian geometry for eeg-based brain-computer interfaces; a primer and a review. *Brain-Computer Interfaces*, 4(3):155–174, 2017.

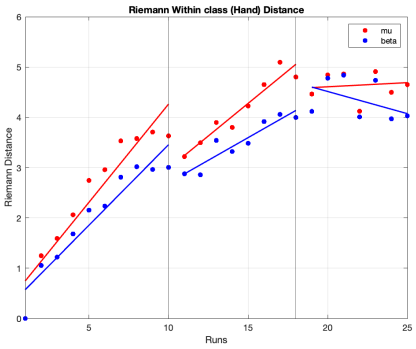
[50] Alexandre Barachant and Stephane Bonnet. Channel selection procedure using riemannian distance for bci applications. In *2011 5th International IEEE/EMBS Conference on Neural Engineering*, pages 348–351. IEEE, 2011.

[51] Wei-Peng Teo and Effie Chew. Is motor-imagery brain-computer interface feasible in stroke rehabilitation? *PM&R*, 6(8):723–728, 2014.

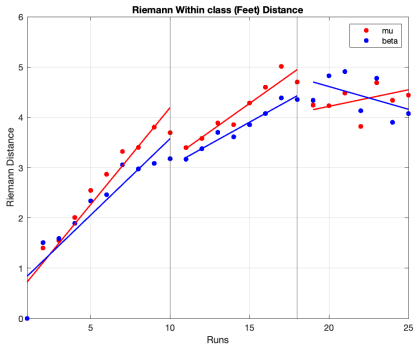
# I. Riemann distances




(a) RD between classes

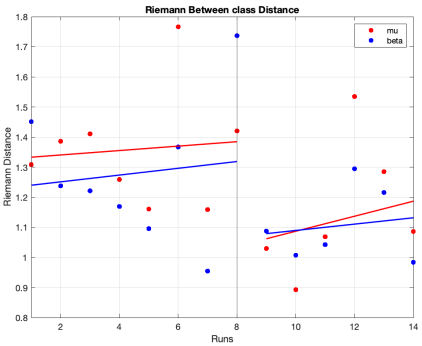



(b) RD within classes (hands)

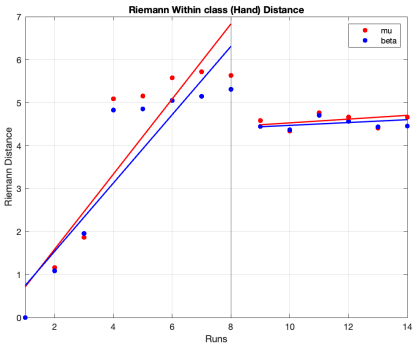



(c) RD within classes (feet)

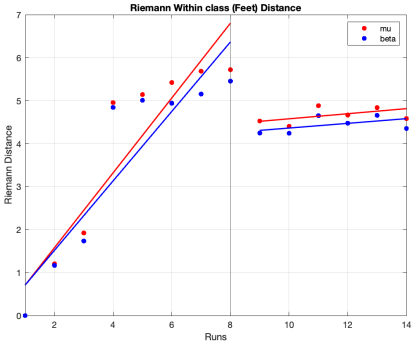
Figure 6.1: Riemann graphs for a8




(a) RD between classes

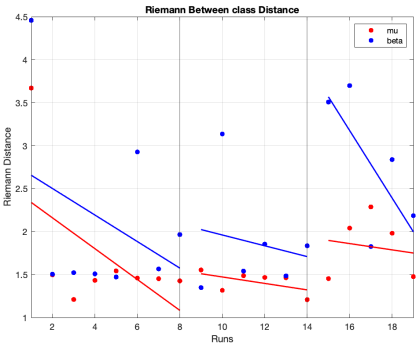



(b) RD within classes (hands)

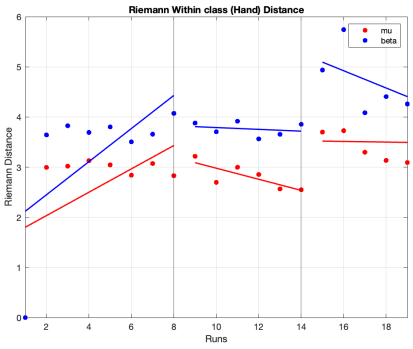



(c) RD within classes (feet)

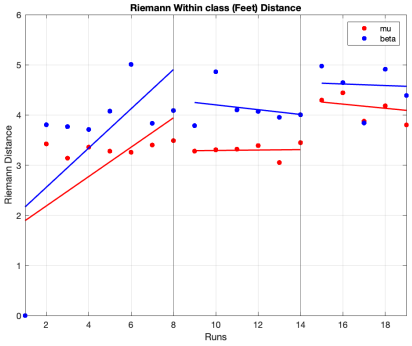
Figure 6.2: Riemann graphs for b9




(a) RD between classes

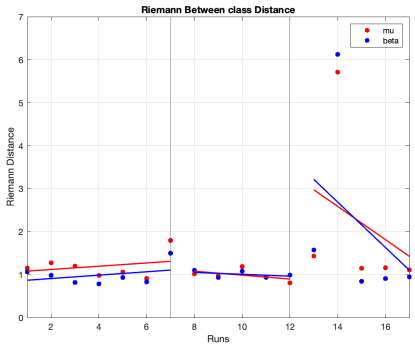



(b) RD within classes (hands)

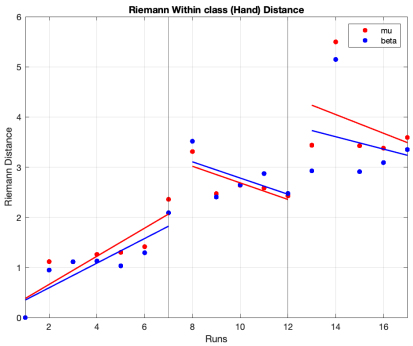



(c) RD within classes (feet)

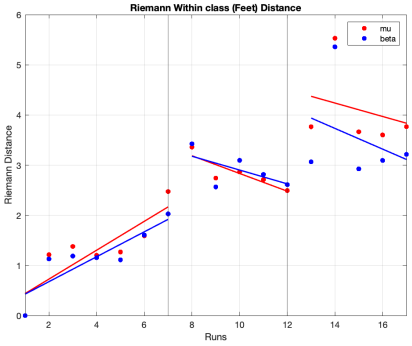
Figure 6.3: Riemann graphs for b1




(a) RD between classes



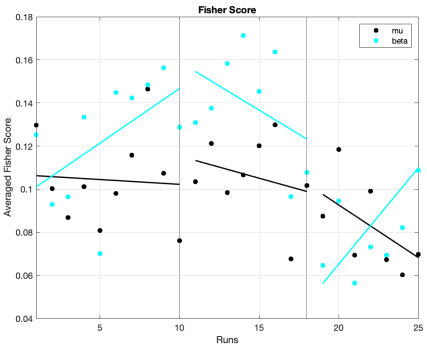

(b) RD within classes (hands)



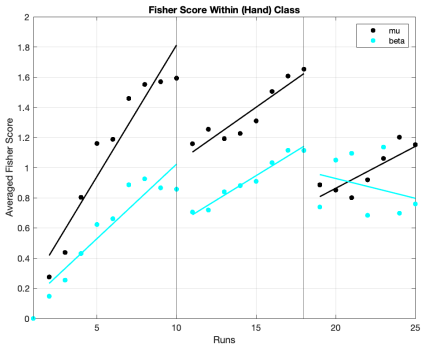

(c) RD within classes (feet)

Figure 6.4: Riemann graphs for a5

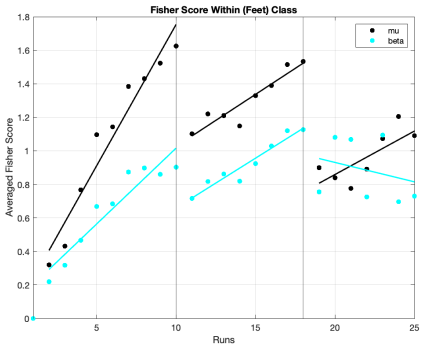
# II. Fisher scores




(a) FS between classes

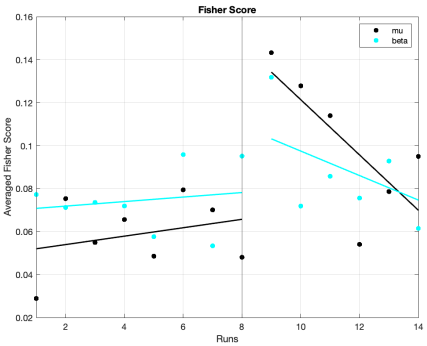



(b) FS within classes (hands)

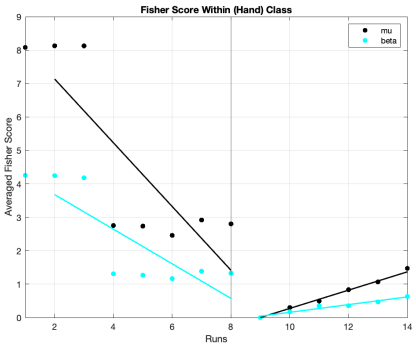



(c) FS within classes (feet)

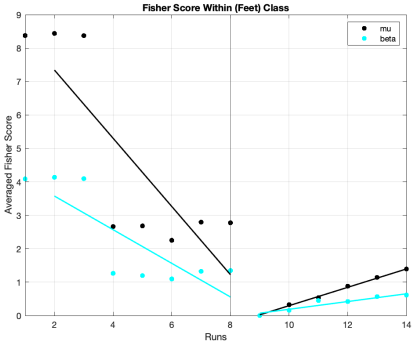
Figure 6.5: Fisher graphs for a8




(a) FS between classes

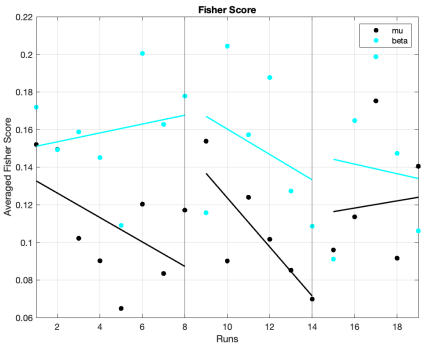



(b) FS within classes (hands)

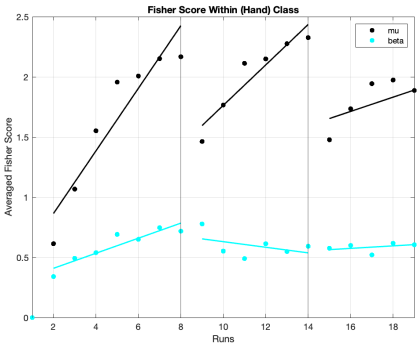



(c) FS within classes (feet)

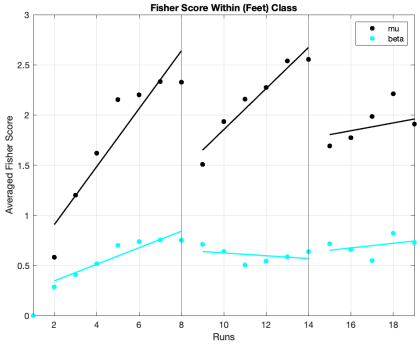
Figure 6.6: Fisher graphs for b9




(a) FS between classes

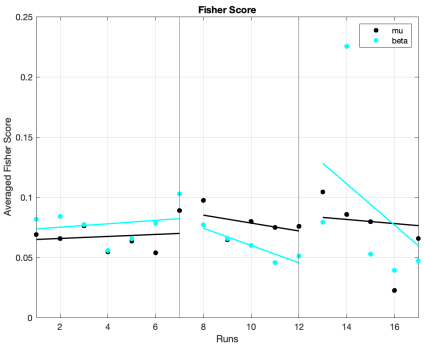



(b) FS within classes (hands)

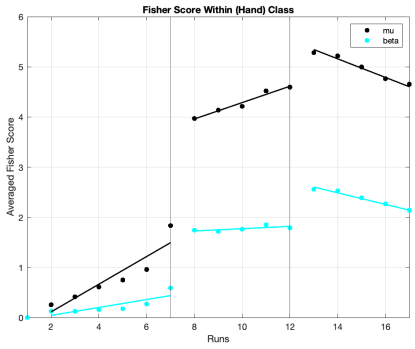



(c) FS within classes (feet)

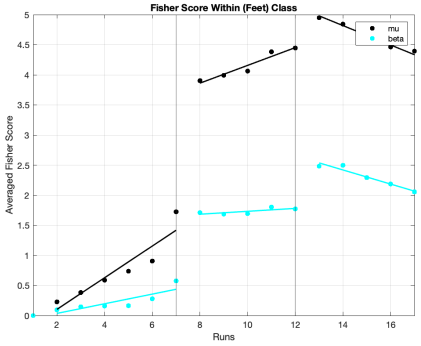
Figure 6.7: Fisher graphs for b1




(a) FS between classes




(b) FS within classes (hands)




(c) FS within classes (feet)

Figure 6.8: Riemann graphs for a5

# III. ERD/ERS

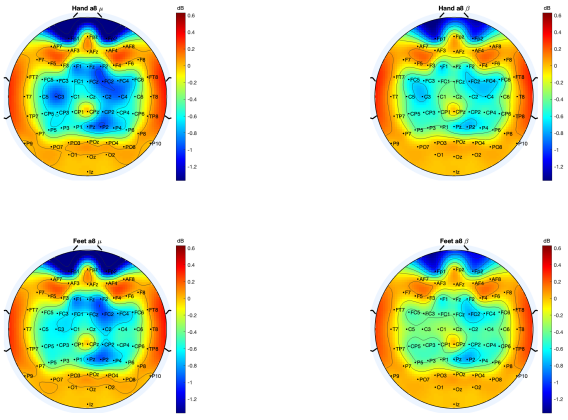


Figure 6.9: ERD/ERS topoplot for a8

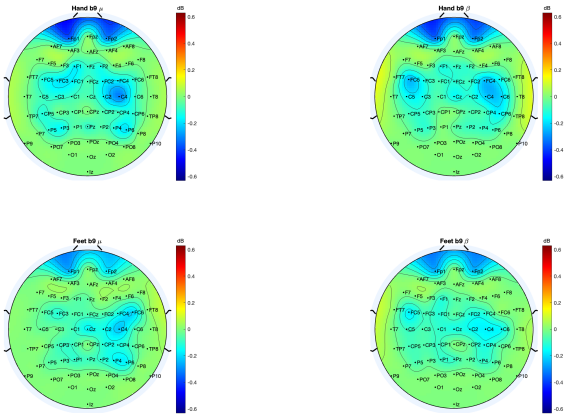


Figure 6.10: ERD/ERS topoplot for b9

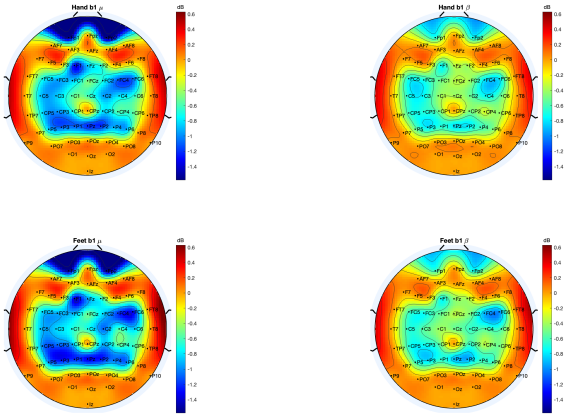


Figure 6.11: ERD/ERS topoplot for b1

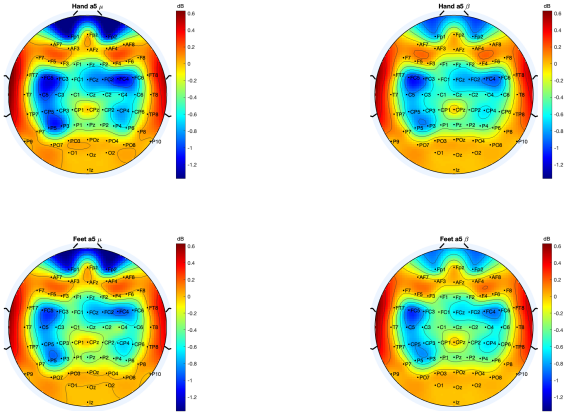


Figure 6.12: ERD/ERS topoplot for a5

# IV. Spectrograms

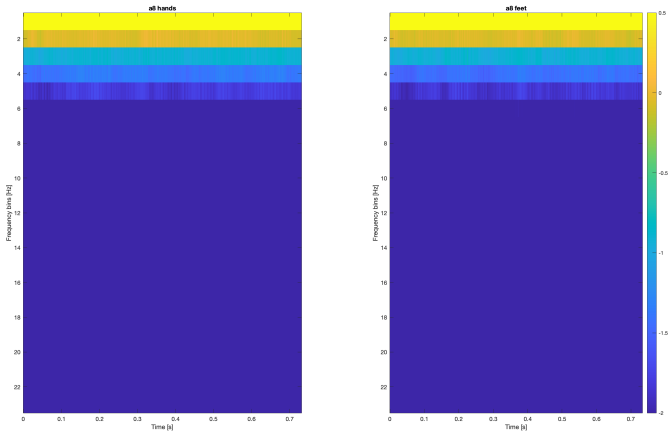



Figure 6.13: Spectrograms for a8

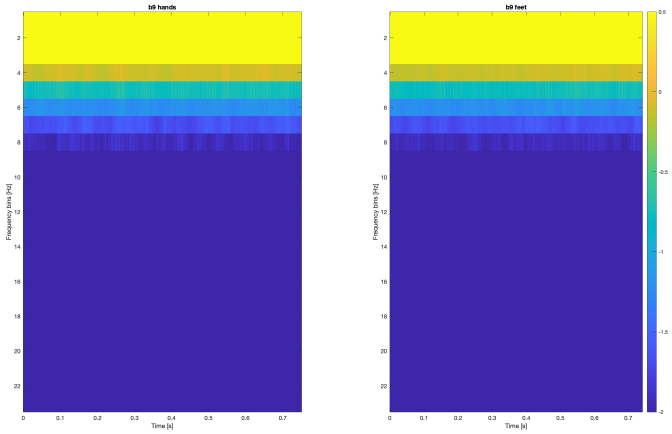



Figure 6.14: Spectrograms for b9

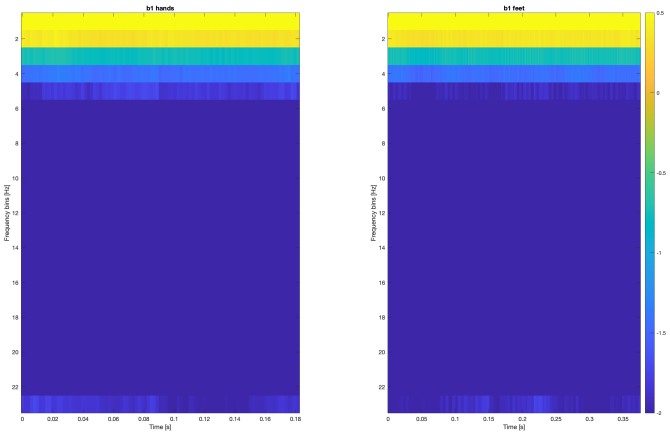



Figure 6.15: Sepctrograms for b1

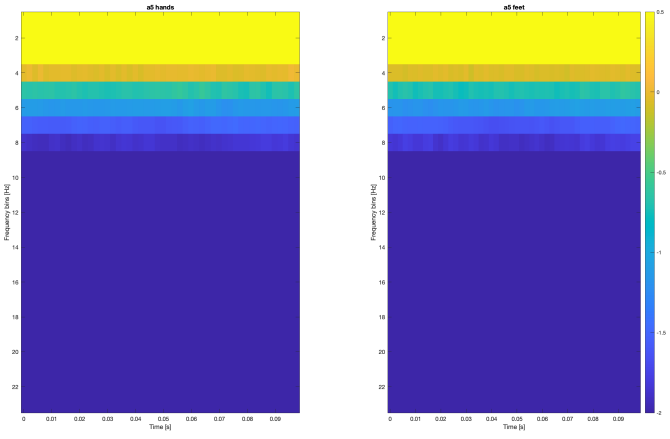



Figure 6.16: Spectrograms for a5

# V. Calibration-Fisher maps

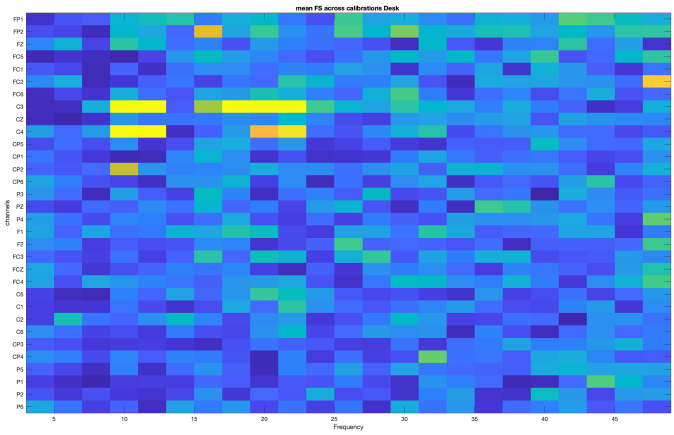



Figure 6.17: Calibration map for a8

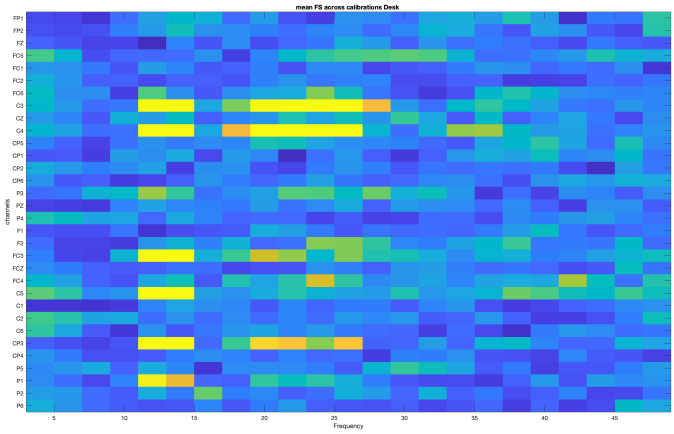



Figure 6.18: Calibration map for b1

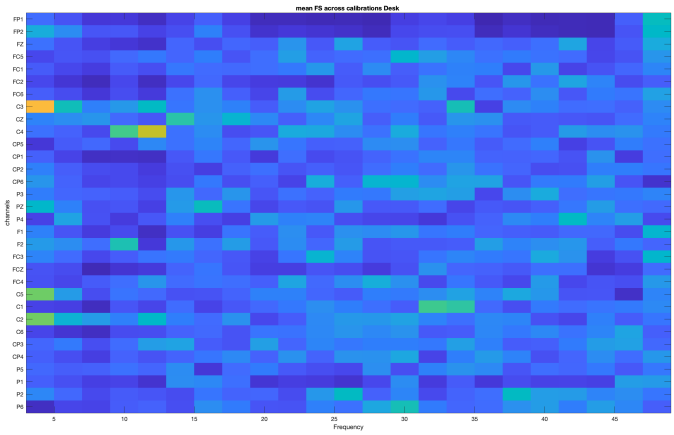
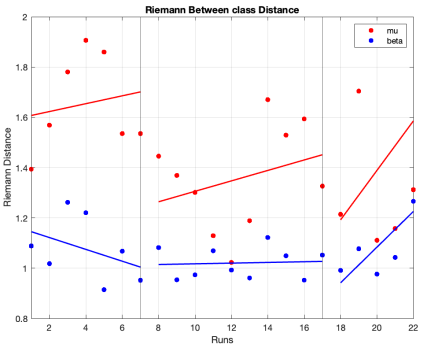
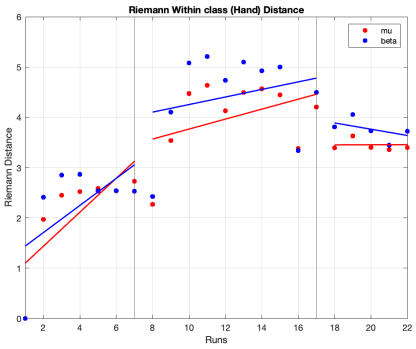



Figure 6.19: Calibration map for a5

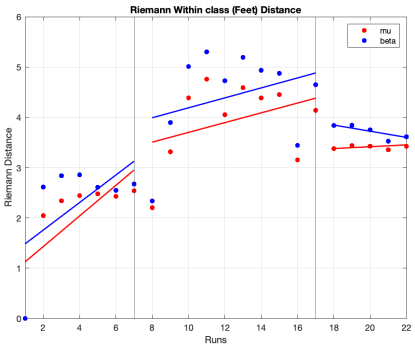
# VI. Riemann distances




(a) RD between classes

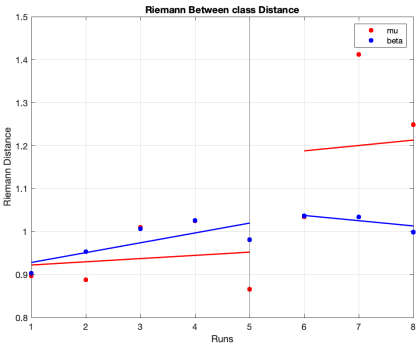



(b) RD within classes (hands)

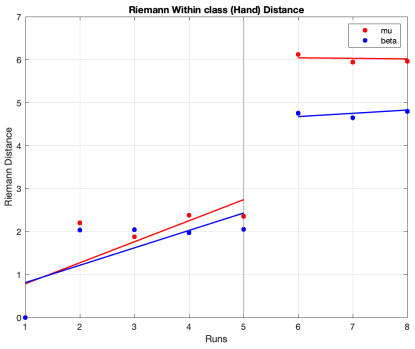



(c) RD within classes (feet)

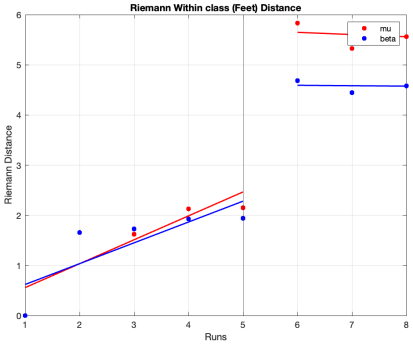
Figure 6.20: Riemann graphs for c1




(a) RD between classes

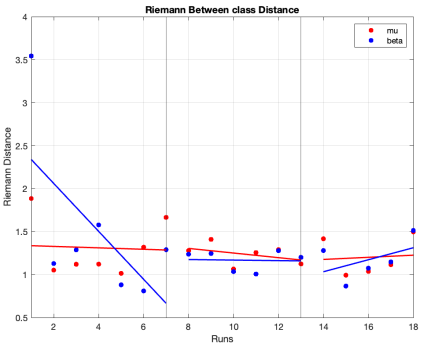



(b) RD within classes (hands)

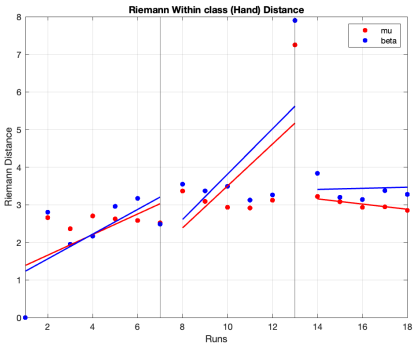



(c) RD within classes (feet)

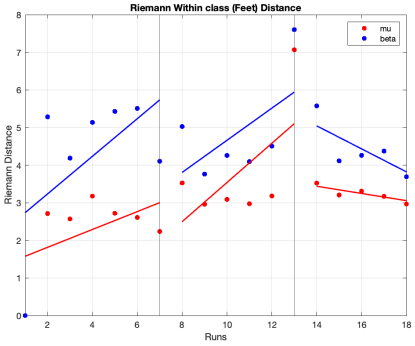
Figure 6.21: Riemann graphs for a3




(a) RD between classes

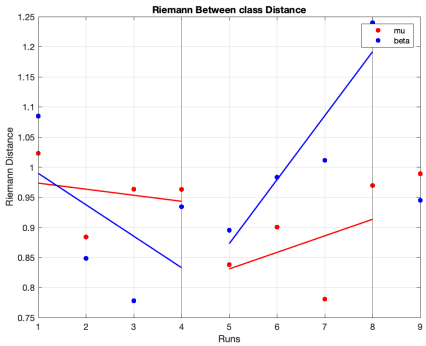



(b) RD within classes (hands)

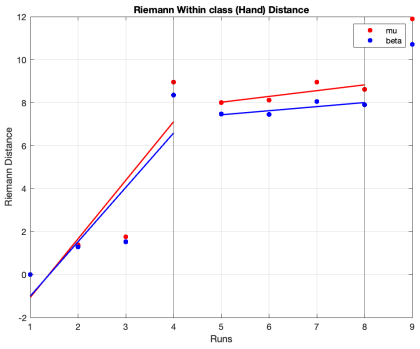



(c) RD within classes (feet)

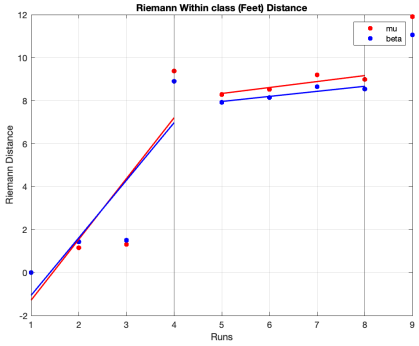
Figure 6.22: Riemann graphs for c4




(a) RD between classes



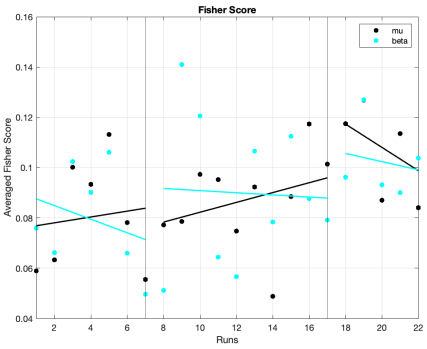

(b) RD within classes (hands)



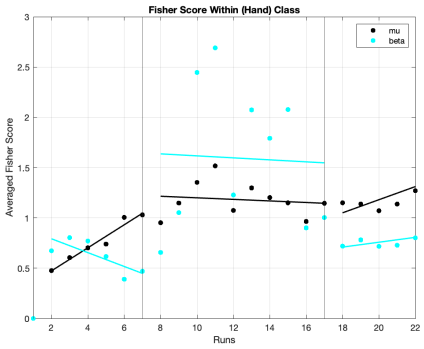

(c) RD within classes (feet)

Figure 6.23: Riemann graphs for c6

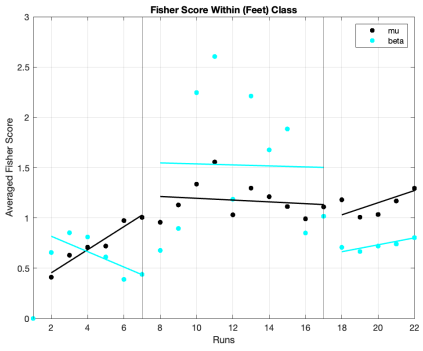
# VII. Fisher scores




(a) FS between classes

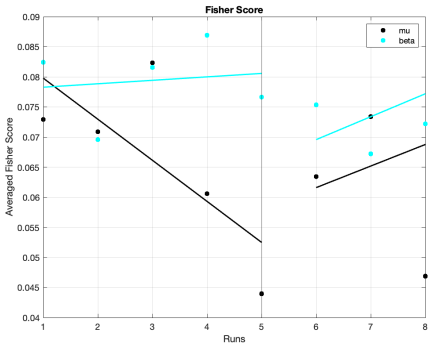



(b) FS within classes (hands)

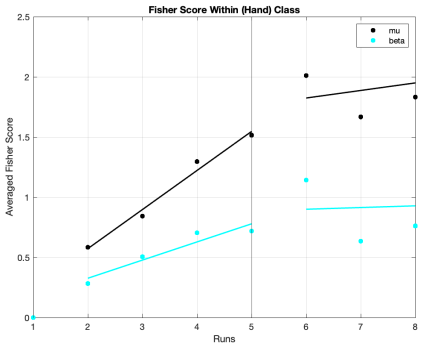



(c) FS within classes (feet)

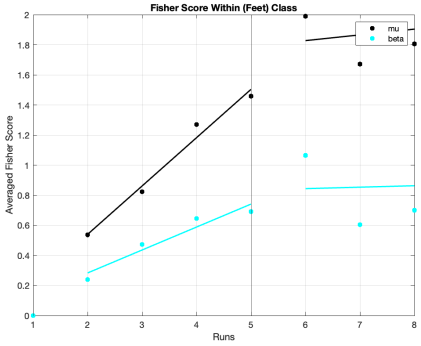
Figure 6.24: Fisher graphs for c1




(a) FS between classes

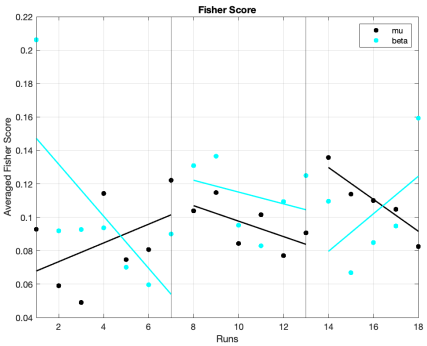



(b) FS within classes (hands)

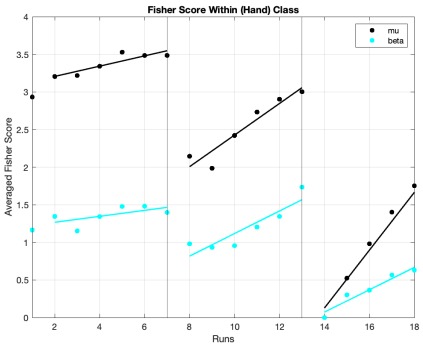



(c) FS within classes (feet)

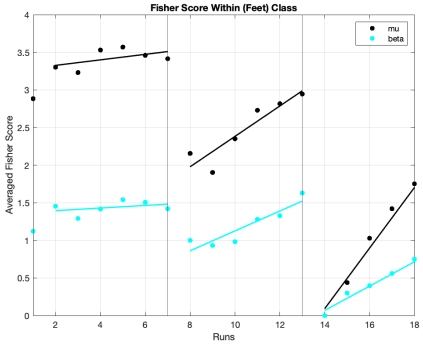
Figure 6.25: Fisher graphs for a3




(a) FS between classes

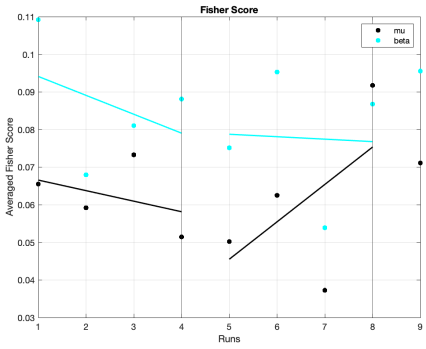



(b) FS within classes (hands)

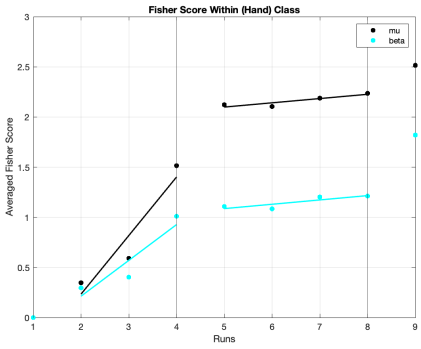



(c) FS within classes (feet)

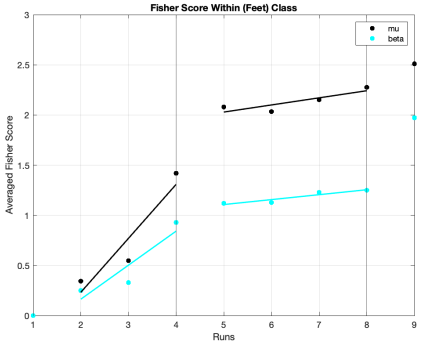
Figure 6.26: Fisher graphs for c4




(a) FS between classes




(b) FS within classes (hands)




(c) FS within classes (feet)

Figure 6.27: Riemann graphs for c6

# VIII. ERD/ERS

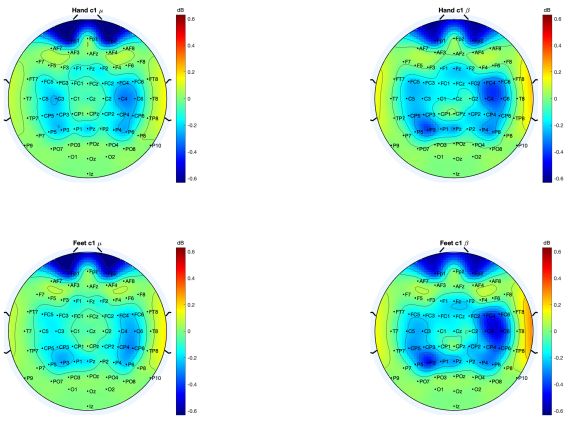


Figure 6.28: ERD/ERS topoplot for c1

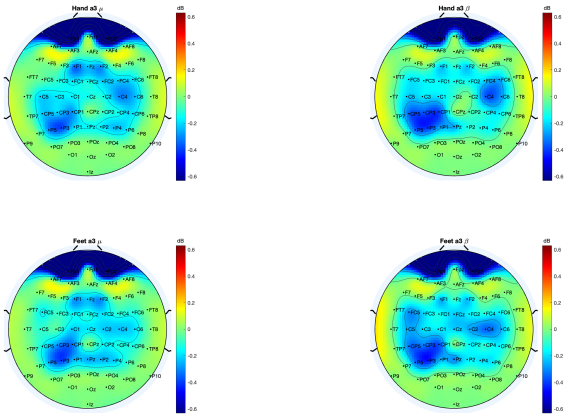


Figure 6.29: ERD/ERS topoplot for a3

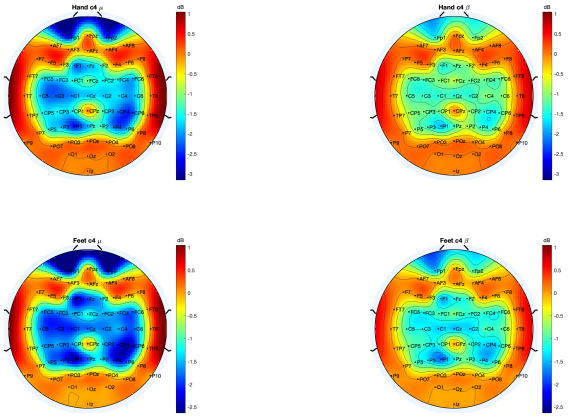


Figure 6.30: ERD/ERS topoplot for c4

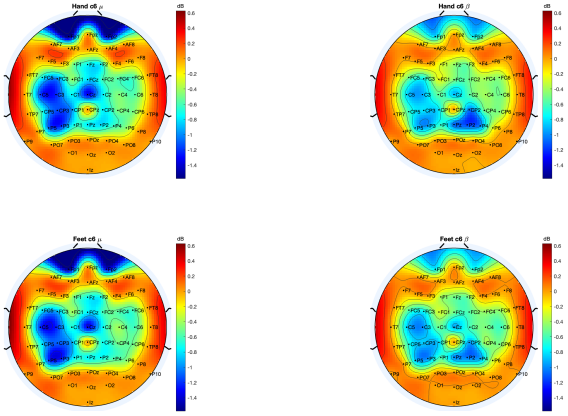


Figure 6.31: ERD/ERS topoplot for c6

# IX. Spectrograms

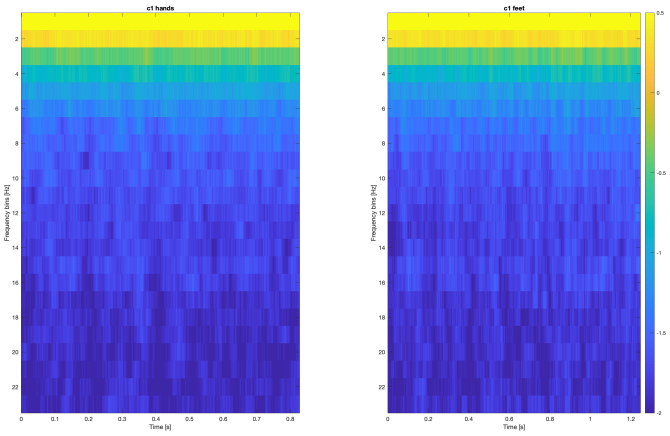



Figure 6.32: Spectrograms for c1

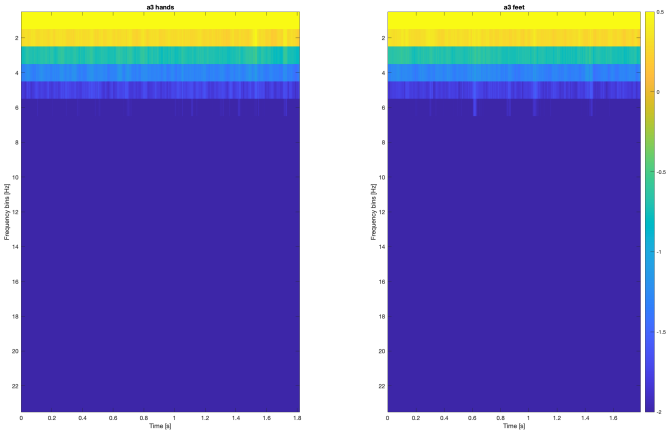



Figure 6.33: Spectrograms for a3

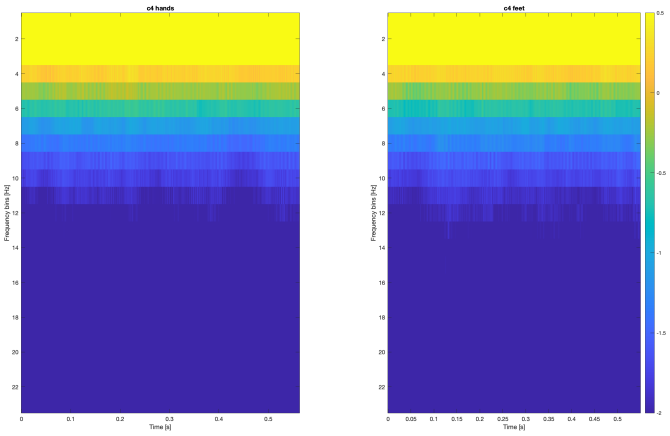



Figure 6.34: Spectrograms for c4

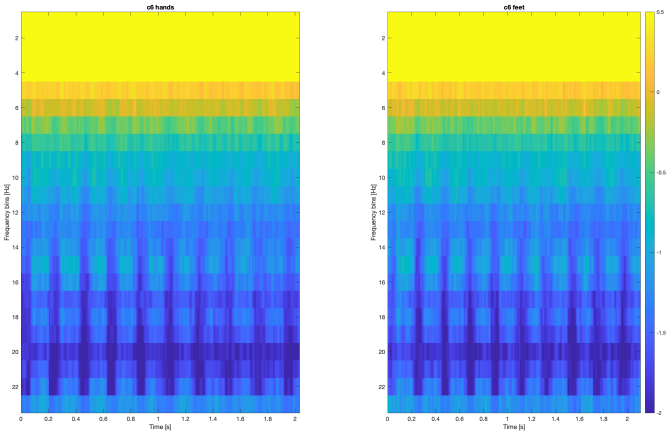



Figure 6.35: Spectrograms for c6

# X. Calibration-Fisher maps

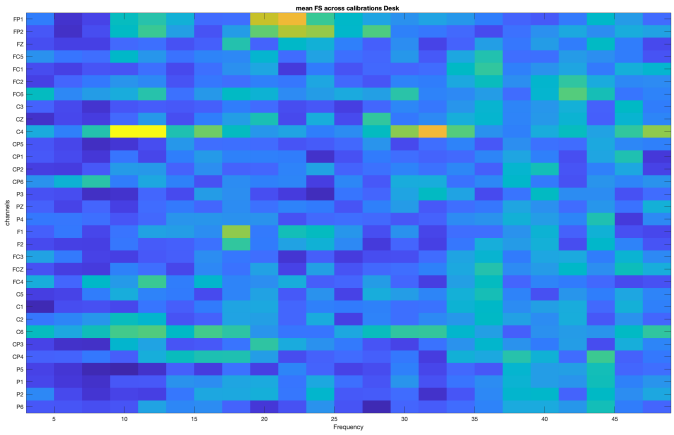



Figure 6.36: Calibration map for c1

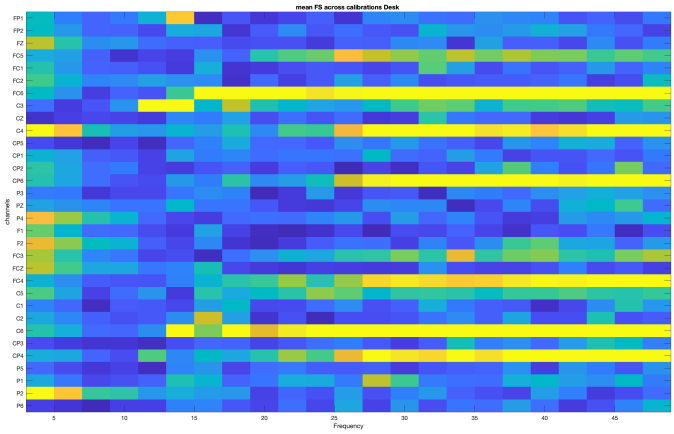



Figure 6.37: Calibration map for c4