Doctoral Dissertation
Doctoral Program in Metrology (37th cycle)

Measurement of motor imagery brain activity for wearable neural interfaces

Applied metrology for wearable neural interfaces

Angela Natalizio

* * * * * *

Supervisor
Prof. Pasquale Arpaia

Doctoral Examination Committee:
Prof. Mauro D'Arco (Reviewer), University of Naples Federico II
Prof. Andrea Cataldo (Reviewer), Università del Salento
Prof. Sabrina Grassini (President), Polytechnic University of Turin
Prof. Luca Lombardo, Polytechnic University of Turin
Prof. Lorenzo Scalise, Marche Polytechnic University

Politecnico di Torino
2025

I hereby declare that, the contents and organisation of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

. . . .

Angela Natalizio
Turin, 2025

Summary

The growing demand for reliable and reproducible electroencephalographic measurements has led to substantial advances in acquisition technologies and signal processing methodologies in both research and clinical contexts. This also applies to real-time brain-computer interface applications. The transition from wired to wireless electroencephalographic systems offers flexibility, wearability and usability. However, this transition introduces metrological challenges, particularly concerning acquisition delay and signal integrity. To ensure the accuracy, repeatability, and reproducibility of analyses based on electroencephalographic signals, a metrological approach is required to quantify and mitigate the sources of uncertainty.

This doctoral thesis first explores the delays in signal acquisition that occur with wireless electroencephalographic devices and assesses how these delays influence the identification of neural phenomena that are time-locked to specific events, such as the P300 component and movement-related cortical potentials. A measurement system was developed to investigate acquisition delays and their associated uncertainties. This system enabled the assessment of signal acquisition delays by measuring the time difference between the start and stop of a reference signal. The goal of this study was to assess the consistency and reliability of various devices. Furthermore, experiments with brain-computer interfaces were conducted to explore how acquisition delays affect the detection of time-locked neural responses. Statistical analyses revealed significant differences in acquisition delays depending on the device and its configuration, with delays ranging from a few tens of milliseconds to over one hundred milliseconds. These variations had a direct impact on the identification of neural phenomena, potentially leading to classification errors. However, compensating for the measured acquisition delay yielded promising results, particularly in enhancing the estimation of P300 latency when using cost-effective recording instruments.

Methods to enhance the reliability of electroencephalographic data were also explored. A comparison was conducted between different techniques for removing unwanted signal components, focusing on configurations that use a limited number of recording electrodes. A combined approach is proposed to improve the removal of unwanted components from electroencephalographic signals. This approach merges artifact subspace reconstruction with multivariate empirical mode decomposition. It proves particularly beneficial when only a few sensors are available, a scenario where traditional techniques are often ineffective. The approach was tested using two publicly available datasets: the first consisting of semi-synthetic data and the second containing experimental recordings. The analysis examined cases with one to four recording electrodes and compared the proposed method to artifact subspace reconstruction alone. Notably, the latter could not be applied when only one electrode was available, whereas the combined approach made this possible. The proposed approach effectively minimized artifacts caused by muscle activity, eye movements, and blinking in both semi-synthetic and experimental data. Interestingly, artifact subspace reconstruction alone performed similarly well on semi-synthetic data. However, when applied to real experimental recordings, it failed to operate correctly.

Then, methods for processing electroencephalographic signals in brain-computer interfaces that focus on imagined movements were examined. A metrological analysis of the literature was conducted to assess advancements in signal classification, pinpoint emerging trends, and tackle the challenges of achieving consistent outcomes. Special emphasis was placed on performance evaluation through a detailed metrological analysis, tacking into account factors like accuracy, uncertainty, repeatability, and reproducibility. The review included 89 studies. The findings indicated a growing dependence on brain-inspired computational models, which have shown to be particularly effective in distinguishing between different movement classes. However, the lack of standardized protocols created difficulties in assessing reproducibility.

In the context of motor imagery-based neural interfaces, it is important to monitor user's level of engagement. This is closely related to the actual generation of motor imagery-related brain activity. Hence, this thesis also discusses how electroencephalography can be used to assess cognitive engagement in real time. A novel approach was created to categorize engagement levels across various tasks by analysing brain activity recordings. The research included twenty-three participants who took part in a modified attention test aimed at eliciting engagement. A classification model was developed to differentiate between engaged and resting states using signals acquired from eight electrodes. The model achieved an average classification accuracy of 90 % on independent test data.

Building on the insights from earlier chapters, the final chapter of this work focuses on the development and experimental validation of a wearable brain-computer interface that enables real-time control of a video game through self-paced motor imagery. Electroencephalographic signals were acquired using eight wet electrodes. A processing pipeline was established to classify three mental tasks: imagining the movement of the left hand, the right hand, or remaining at rest. This classification allowed for real-time control of an avatar in an endless runner game. The study involved twenty-three participants, and a custom metric was created to assess performance during gameplay. The mean classification accuracies for differentiating between left hand vs. rest, right hand vs. rest, and left hand vs. right hand were 73 %, 73 %, and 67 %, respectively. Participants who demonstrated higher classification accuracy during offline calibration exhibited better control of the avatar in the game. This research contributes to the development of wearable, self-paced neural interfaces for real-time applications, facilitating more natural and intuitive interactions compared to traditional synchronous systems. By moving beyond benchmark datasets and testing in an interactive gaming environment, this work lays the groundwork for future applications in mobile and everyday-use brain-computer interfaces.

Acknowledgements

As I reach the conclusion of this remarkable journey, I want to take a moment to thank the many people who have been essential to my progress.

First and foremost, my deepest gratitude goes to Prof. Pasquale Arpaia for his invaluable guidance, support, and encouragement throughout my journey, even before I began my PhD. He has always been a pillar of support, helping me to grow both professionally and personally. A pragmatic and warm person with a profound passion for metrology, he has instilled in me the importance of being practical in a world that often gets lost in words. His expertise and mentorship have played a crucial role in shaping my research and professional growth, for which I will always be thankful.

A special thank you goes to Prof. Marco Parvis, who has always been present, even as he began facing challenging times. His dedication and deep passion for his work have truly inspired us all. No matter the circumstances, he has always been ready to offer guidance and has shown a remarkable commitment both to his students and to the field of metrology.

I would like to sincerely thank Prof. Sabrina Grassini, whose kindness and constant support have been invaluable. Her careful guidance has truly marked my academic journey.

I am truly grateful for the opportunity to meet and collaborate with Dr. Christoph Guger. I want to express my gratitude for his warm welcome during my international research experience at g.tec medical engineering GmbH. His hospitality and the opportunity to work in such an inspiring environment have greatly contributed to my professional and personal development. I would also like to sincerely thank everyone who made my time in Austria truly memorable.

I am also deeply grateful to Antonio Esposito, whom at some point I should have started calling “professor”. He has been more than just a colleague; he has been a genuine companion throughout this experience. His support, expertise, and friendship have made this journey much more meaningful and fulfilling.

I want to express my sincere gratitude to all the incredible members of ARHeMLab – Augmented Reality for Health Monitoring Laboratory. Being part of this wonderful group has been a true privilege, and I appreciate the engaging discussions, collaborations, and friendships that have greatly enhanced my experience. A special mention goes to Fortuna and Enza, who have been amazing companions—there wasn’t a day that went by without a smile thanks to them.

A special thank you to Isa, Leila, Leo, Amina, and everyone else who has been so welcoming and ready to help whenever I needed it. Their kindness and generosity have greatly impacted my journey, and I am truly grateful to have met them.

Finally, I want to thank all the amazing people I have met throughout these years at conferences and PhD programs around the world. Every conversation, every shared idea, and every new perspective have played a significant role in my growth.

To everyone who has stood by me and had faith in me throughout this journey: thank you. This achievement would not have been achievable without your support.

I would like to dedicate this thesis to my beloved family

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Over the past three decades, the field of human-machine interaction has undergone significant advancements. This growth has led to increased scientific interest in using biomedical data for various applications, such as communication, environmental monitoring and neurorehabilitation. One of the main goals of biomedical data analysis, especially in signal processing, is the development of intuitive human-machine interfaces (HMIs) [1]. Brain-computer interfaces (BCIs) are an important subset of HMIs and will be discussed in detail in this thesis work.

BCIs are a revolutionary technology that creates a direct link between the brain and external devices, bypassing traditional muscle interactions [2]. Since their inception, various technologies have been developed to capture brain signals that act as inputs for BCIs [3]. The algorithms used in these systems have progressed from simple analyses of the temporal, frequency, and spatial characteristics of brain signals to sophisticated deep learning techniques. These advancements have enabled BCIs to find applications in a wide range of fields, from healthcare to entertainment. Furthermore, they have prompted the BCI community to start a conversation in early 2024 focused on refining the definition of BCIs [4].

This chapter provides essential background information to introduce the reader to these systems and to facilitate the understanding of subsequent chapters.

1.1 Definition and purpose of brain-computer interfaces

One of the most widely used definitions of BCI was published by John and Elizabeth Wolpaw in 2012: “A BCI is a system that measures CNS activity and converts it into artificial output that replaces, restores, enhances, supplements, or improves natural CNS output and thereby changes the ongoing interactions between the CNS and its external or internal environment.” [5].

This definition indicates that a BCI alters how the user’s central nervous system (CNS) interacts with the environment. It also views BCIs as output devices since they transmit information from the brain to the outside world.

Given the significant advancements in both academic and commercial neurotechnology since this definition was introduced, it has prompted various questions and highlighted some limitations. Notably, the BCI society initiated a discussion to revise the BCI definition in early 2024 [4].

First, the BCI society questioned whether the definition of BCI should specifically cover the brain or the entire CNS, which also includes the spinal cord. Secondly, according to Wolpaw's definition, a device that merely monitors brain signals does not qualify as BCI because it recognises "interaction" as a fundamental component of the system. However, interactions between the CNS and the environment are almost always bidirectional. This has led to a discussion as to whether the definition of a BCI should describe a system that modulates or modifies brain function in addition to reading and decoding brain signals.

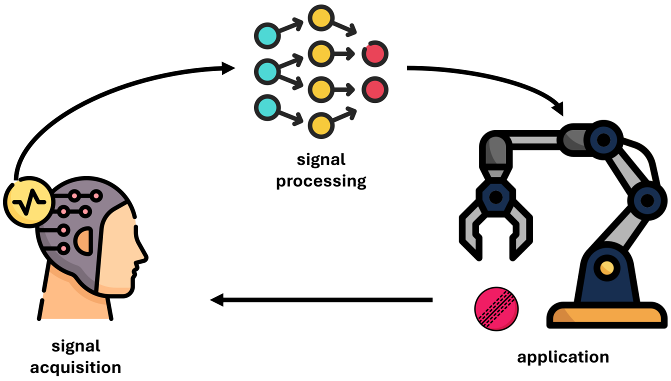


Figure 1.1: The main components of a BCI system.

A new definition was developed based on feedback from 147 BCI stakeholders who participated in a questionnaire designed to refine the existing definition. The updated definition states: "A brain-computer interface is a system that measures brain activity and converts it in (nearly) real-time into functionally useful outputs to replace, restore, enhance, supplement, and/or improve the natural outputs of the brain, thereby changing the ongoing interactions between the brain and its external or internal environments. It may additionally modify brain activity using targeted delivery of stimuli to create functionally useful inputs to the brain." [4]. This revised definition captures both the measurement and modulation aspects of BCIs, offering a more thorough understanding of their capabilities and potential applications.

Fig. 1.1 shows the main components of a BCI system that are: the acquisition of the brain signal, the processing of the acquired signal, and a final application/feedback that can be either naturally or artificially provided [3].

1.1.1 Invasiveness-based classification

Based on the degree of invasiveness of the techniques used to acquire brain signals, BCI systems can be classified into invasive, semi-invasive and non-invasive categories [6]. Each category of BCI systems offer distinct advantages and challenges, influenced mainly by trade-offs in terms of spatial and temporal resolution, security and practicality [7].

Invasive BCI involves implanting electrodes directly into the brain tissue, offering maximum spatial resolution [8]. They are able to capture neuronal activity at the level of single neurons or small groups of neurons. Typically, microelectrode arrays and depth electrodes are implanted, offering the highest temporal (around 0.003 s) and spatial (down to 0.05 mm) resolution. However, the invasiveness of these techniques carries significant risks, including infection, inflammation and long-term biocompatibility issues. The surgical procedures required for electrode implantation are complex and carry inherent medical

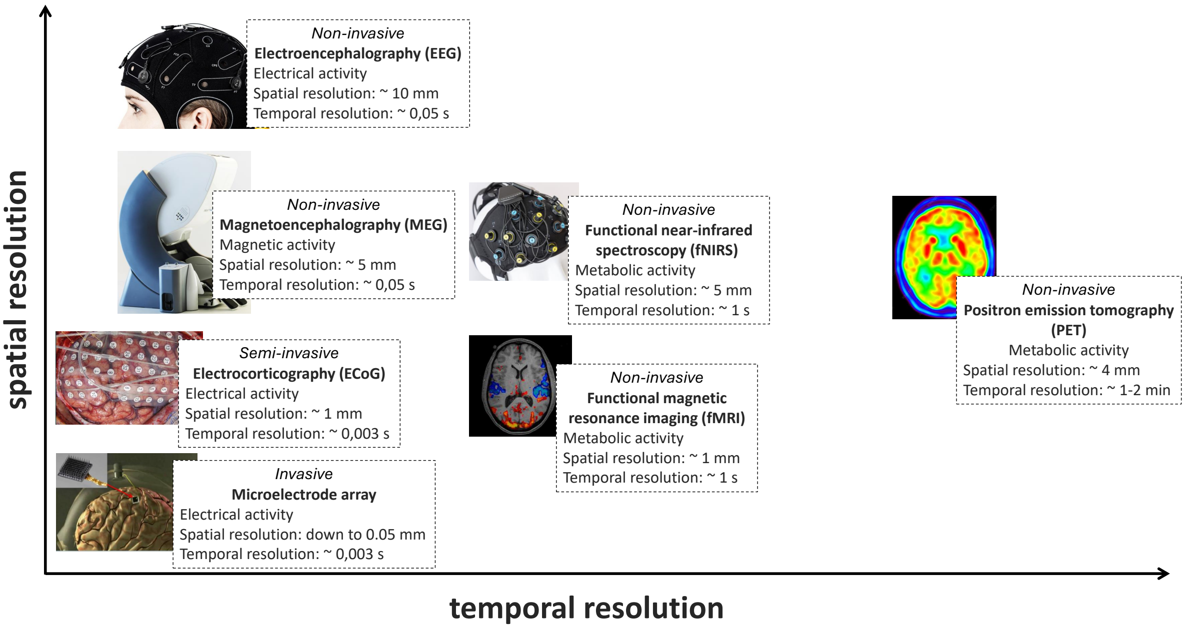


Figure 1.2: Brain signal acquisition modalities according to their spatial and temporal resolution.

risks, limiting the widespread use of invasive BCI to highly controlled clinical and research environments.

Semi-invasive BCI involves placing electrodes on the brain's surface (electrocorticography, ECoG) or within the skull but outside the brain tissue [9]. During a surgical procedure, ECoG electrodes are positioned directly on the cerebral cortex. This method strikes a balance between invasiveness and data fidelity, making it suitable for specific clinical applications like epilepsy monitoring and research that requires accurate cortical mapping. ECoG provides spatial resolution at the millimeter scale and can record a wide frequency range, reaching up to 200 Hz or more. It also captures neural signals with amplitudes of up to 100 µV, offering a clear and detailed perspective on brain activity. While semi-invasive techniques lower the risks associated with fully invasive methods, they still require surgical procedures and carry potential complications, such as infection or electrode displacement.

Non-invasive BCIs involve the acquisition of brain signals without penetrating the scalp. This makes them less risky and easier to use. The most common non-invasive technique is electroencephalography (EEG), which involves placing electrodes on the scalp to measure the electrical activity generated by the brain [10]. EEG is very popular due to its high temporal resolution, which allows real-time monitoring of brain activity with a temporal resolution of about 0.05 s. Other advantages are its cost-effectiveness, ease of use, wearability and portability, which make it suitable for both clinical and research purposes. However, it offers a spatial resolution of about 10 mm, which is lower than more invasive methods. Magnetoencephalography (MEG) is another non-invasive technique that measures the magnetic fields produced by neural activity [11]. MEG provides better spatial resolution than EEG, at around 5 mm, while maintaining a comparable temporal resolution of approximately 0.05 s. This technique offers a good balance between spatial and temporal resolution but is more expensive and less portable than EEG. Functional near-infrared spectroscopy (fNIRS) is a technique that measures metabolic activity by detecting changes in blood oxygenation levels [12]. fNIRS provides a spatial resolution of about 5 mm and a temporal resolution of around 1 s. It is portable, making it useful for applications where mobility and ease of use are important, although it offers lower temporal resolution compared to EEG and MEG. Functional magnetic resonance imaging (fMRI) measures brain activity by detecting changes in blood flow, reflecting metabolic activity [13]. Like ECoG, it provides excellent spatial resolution of about 1 mm, making it highly suitable for detailed brain mapping. However, its temporal resolution is relatively poor, at around 1 s, due to the slow nature of hemodynamic responses. fMRI is widely used in research but is less practical for real-time applications due to its temporal limitations and the need for large, stationary equipment. Positron emission tomography (PET) is another example non-invasive technique that measures metabolic activity by detecting radioactive tracers injected into the bloodstream [14]. PET offers a spatial resolution of approximately 4 mm but has a very poor temporal resolution, ranging from 1 min to 2 min. PET is also characterised by its large, expensive, and unhandy equipment, making it less practical for routine or real-time applications [1]. This makes PET suitable for studying long-term metabolic changes in the brain rather than real-time monitoring.

Fig. 1.2 summarizes the various techniques for acquiring brain signals, highlighting their spatial and temporal resolutions. It provides a visual comparison of their capabilities and limitations.

1.1.2 Paradigm-based classification

BCIs are also classified based on the type of paradigm used, namely active, reactive, and passive interfaces [15, 1].

Active BCIs generate outputs directly from voluntary brain activity, in which users intentionally modulate their brain signals to control an external device. These BCIs are based on spontaneous signals, which are not triggered by external events. A common example of this paradigm are the motor imagery-based BCIs [16, 17]. Users imagine specific movements of body parts without performing them in order to generate corresponding neural signals that the BCI system interprets as control commands. While their main strength lies in their independence from any external stimulus and thus everything depends on the user's imaginative capacity, this is also a disadvantage. Indeed, these types of BCIs require significant mental effort and training on the part of the user. Furthermore, performance can be affected by fatigue or lack of concentration on the part of the user [18].

Reactive BCIs are based on evoked signals. They respond to specific stimuli presented to the user by detecting automatic brain responses to generate control signals [19]. For example, a reactive BCI can use visual, haptic or auditory stimuli to elicit specific brain responses, such as event-related potentials (ERPs). They have a reduced cognitive load compared to active BCIs, as the system exploits the brain's natural and automatic responses. In addition, they have faster response times and lower training requirements than active BCIs. However, reactive BCIs face challenges related to dependence on external stimuli, limited control versatility and potential user distraction or fatigue, especially in the long-term use scenario [15].

Passive BCIs enable the investigation of various cognitive states, including fatigue, vigilance, workload, emotions, and engagement [20, 21, 22]. Passive BCIs monitor the user's brain activity without any active effort or specific responses to stimuli, as there is no need for intentional control or response [23]. These systems are designed to infer the user's mental state, such as attention levels, engagement, workload or emotional states. Furthermore, they are suitable for adaptive systems that adjust their behaviour according to the user's mental state, enhancing the user's experience and safety. However, they have less direct control over external devices, as the system interprets rather than responds directly to the user's intentions. Furthermore, there are potential privacy and ethical issues, as the continuous monitoring of mental states may reveal sensitive information about the user [24].

1.2 Electroencephalography-based systems

As mentioned in the previous section, EEG is one of the most popular and widely used techniques for brain signal acquisition in the field of BCIs. The origins of the EEG date back to the beginning of the 20th century, when the German neuropsychiatrist Hans Berger made the revolutionary discovery of electrical activity in the human brain [25]. In 1924, Berger recorded the first human EEG, thus pioneering the field of electrophysiology [1].

The research conducted by Berger demonstrated that it is possible to measure the electrical activity of the brain non-invasively by placing electrodes on the scalp [26]. This discovery laid the foundation for numerous advances in neuroscience and neurotechnology, leading to the development of modern EEG systems. Today, EEG is valued for its non-invasive nature, portability and relatively low cost, making it accessible and practical for a wide range of applications. These advantages make EEG particularly suitable for real-world applications, extending beyond clinical settings to everyday use. A summary of the strengths and limitations of EEG in relation to BCIs can be found in Table 1.1.

Table 1.1: Strengths and limitations of EEG technology in BCIs

Strengths	**Limitations**
EEG is a non-invasive method for recording brain activity, ensuring safety and comfort for users.	EEG has lower spatial resolution compared to other imaging methods such as fMRI, which makes it difficult to pinpoint the exact areas of brain activity.
EEG has good temporal resolution, allowing for real-time capture of brain activity with millisecond accuracy.	The signals obtained from EEG can be affected by external noise and artifacts, such as eye blinks and muscle movements.
EEG is relatively inexpensive compared to other neuroimaging techniques.	The signals are also inherently non-stationary, resulting in considerable variability both within and between individuals.
Setting up and using EEG systems is generally straightforward, making them accessible for a variety of applications.	the low signal-to-noise ratio of EEG signals requires careful preprocessing, which can affect the reliability of the recorded data.
EEG equipment is often portable, providing flexibility for use in different settings.	primarily captures cortical activity, which means it may not effectively measure deeper brain structures.

In this thesis, the focus will be solely on EEG as the method for acquiring brain signals, highlighting its essential role in creating effective and reliable BCIs.

1.2.1 Data acquisition

The accuracy and reliability of EEG data are highly dependent on the correct positioning of the electrodes on the scalp, the type of electrodes used and the characteristics of the measurement system. This section explores the fundamental aspects of EEG acquisition, highlighting the importance of each component in the overall process.

The so-called "10-20 system" was first proposed by Herbert Jasper in 1958 and is internationally standardized. It is the most well-known and widely used method for electrode placement in EEG acquisition [27, 1]. This system ensures consistency and reproducibility in EEG recordings, facilitating the comparison of data across different studies and clinical settings. In the 10-20 system, the term “10-20” assumes that the distances between adjacent electrodes are 10 % or 20 % of the total front-to-back or right-to-left distance of the skull.

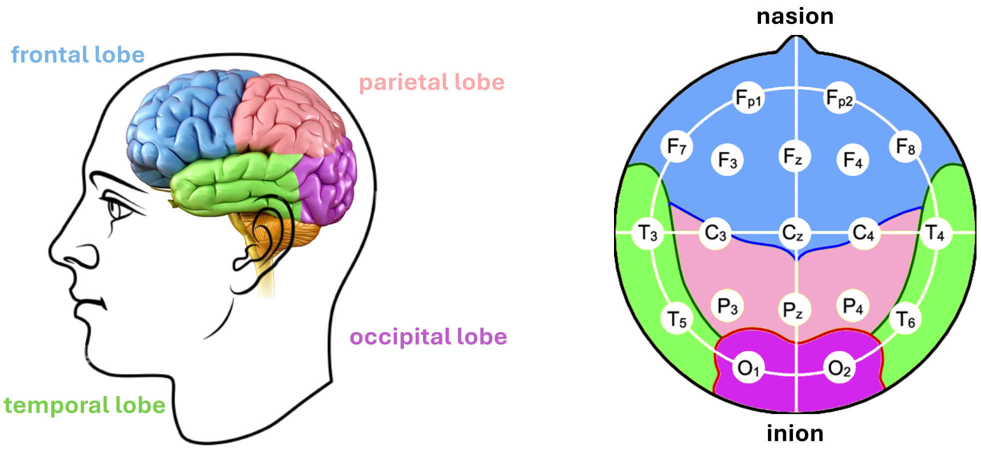


Figure 1.3: Visual representation of brain regions (on the left) and of the 10-20 system for electrode placement (on the right).

Initially designed to place a relatively small number of electrodes (approximately 21), the 10-20 system has evolved with the advent of multi-channel EEG technology. This evolution led to the development of the 10-10 system (approximately 81 electrodes) and the 10-5 system (approximately 320 electrodes).

The system uses specific anatomical points such as the nasion (the junction between the forehead and nose) and the inion (the bony prominence at the back of the skull) as reference points. Electrodes are placed in predetermined positions labeled with letters and numbers, where letters indicate brain regions (e.g., F for frontal, T for temporal, P for parietal, O for occipital) and numbers indicate hemispheres (odd numbers for the left hemisphere, even numbers for the right hemisphere). The letter “z” denotes electrodes placed on the midline, i.e., the central sagittal reference curve. This standardised approach allows systematic coverage of the entire scalp, ensuring that all major brain regions are adequately represented in the EEG recordings.

Fig. 1.3 illustrates the brain regions on the left and the electrode placement using the 10-20 system on the right.

The type of electrode used in EEG acquisition significantly influences the quality of the recorded signals. Traditionally, wet electrodes have been the most commonly used electrode type due to their excellent conductivity, excellent signal-to-noise ratio and high reliability [28]. These electrodes require the application of a conductive gel or paste between the electrode and the scalp to reduce impedance and ensure optimal electrical contact. Despite the provision of high-quality signals, the application process is often time-consuming and typically necessitates the expertise of a trained professional to apply the conductive material. Furthermore, extended usage of wet electrodes may result in discomfort or dermatological irritation, limiting their practical application in non-clinical and non-laboratory contexts.

To address some of the limitations associated with wet electrodes, semi-dry and dry electrodes have been developed. Semi-dry electrodes use a saline solution instead of gel, which is slowly released to maintain a good electrical connection with the scalp

[29]. This design reduces preparation time and improves user comfort, making it more suitable for various applications. However, semi-dry electrodes can face challenges such as the uncontrolled and unexpected release of the saline solution, which can compromise signal stability and reliability. Dry electrodes, which do not require any conductive gel or paste, are typically made from highly conductive materials such as silver or gold. They rely on mechanical pressure to ensure contact with the scalp. Unlike wet and semi-dry electrodes, dry electrodes are often designed with a non-flat shape to penetrate hair and make direct contact with the scalp [30]. While dry electrodes offer the advantage of quick and easy application, they often exhibit higher impedance levels (ranging from a few kΩ for wet electrodes to several hundred kΩ for dry electrodes) and are more susceptible to movement artefacts [28]. These factors can negatively impact signal quality, particularly in dynamic or mobile environments.

Electrodes can be further classified as active or passive according to their internal components. Passive electrodes are simple conductors that capture the electrical activity of the brain and require external amplification. They are more prone to noise and interference, which can degrade the fidelity of recorded signals. In contrast, active electrodes have built-in amplifiers that amplify the signal directly at the point of contact with the scalp. This minimises susceptibility to external noise and electromagnetic interference, improving the signal-to-noise ratio [31]. Active electrodes are particularly useful in environments with strong electromagnetic interference. However, they tend to be more expensive, bulky and complex than passive electrodes, although they offer superior performance in terms of signal integrity and measurement precision.

Once acquired from the electrodes, the EEG signal needs to be amplified, filtered and digitised [32, 33]. Amplification is the first step in the processing chain as the electrical signals generated by the brain are of the order of µV. This step, together with filtering, is essential for conditioning signals prior to voltage acquisition. Filtering serves to isolate the frequency bands of interest, which are generally limited to a range of up to 100 Hz [15]. However, this upper limit can be adjusted according to the specific requirements of the study or application. Effective filtering aims to maximise the signal-to-noise ratio, which is inherently low in EEG measurements. By focusing on relevant frequencies, filtering minimises the impact of noise and unwanted artefacts, thus improving the quality of recorded signals With the advancement of digital electronics, the voltage acquisition process has evolved to include analogue-to-digital conversion. This conversion translates analogue EEG signals into digital form, allowing for more efficient processing in subsequent steps. The digitised signals can then be stored in a digital memory for further analysis and interpretation.

The gold standard of EEG devices consists of wired devices with a large number of wet electrodes. The characteristics of this type of system limit it to clinical settings or research laboratories due to the limited mobility of the user [16]. However, there has been a growing trend towards the development and use of wireless EEG devices [34, 35]. These systems are typically equipped with a small number of electrodes and are designed to improve usability and end-user comfort. Although wireless EEG devices offer significant advantages in terms of mobility and ease of use, they pose some trade-offs. The reduced number of electrodes may result in lower spatial resolution and less information acquired. In addition, wireless transmission can introduce higher latency and a greater risk of data loss, potentially impacting the quality and reliability of acquired signals. Despite these challenges, the convenience and ease of use of wireless EEG devices make them an attractive option for a range of applications, including mobile health monitoring, neurofeedback and everyday BCI use.

1.2.2 Characteristics and processing

EEG analysis involves examining the electrical activity of the brain to identify distinct patterns and features that can be used in various applications, such as BCIs, neurological research, and clinical diagnostics [1].

To effectively use EEG signals in BCI applications, it is essential to understand the underlying rhythms and their characteristics. EEG signals are typically divided into five main frequency bands, each associated with different brain states and functions [10]:

- The *delta* band is characterized by frequencies in the range from 0.5 Hz to 4.0 Hz and amplitudes from 20 µV to 200 µV. These rhythms are the slowest EEG waves and are predominantly present when a subject is deeply asleep. They are also observed in infants and in the case of certain brain diseases.
- *Theta* waves, with frequencies from 4.0 Hz to 8.0 Hz and amplitudes from 20 µV to 100 µV, are slightly faster and are associated with drowsiness, light sleep, and early stages of sleep. Theta rhythms are also linked to meditative and creative states.
- The *alpha* band is in the range from 8.0 Hz to 13.0 Hz and is mainly visible in the occipital and parietal regions. These waves are prominent during wakefulness and states of relaxation, especially with the eyes closed. They decrease with opening the eyes and performing mental tasks. When this EEG band, in the same frequency band, is visible in the central regions, i.e. in the sensory-motor area, this is referred to as the μ band. It is associated with motor activities and is evident when the body is physically at rest but mentally active, particularly in the sensorimotor cortex.
- The *beta* band is defined by frequencies between 13.0 Hz to 30.0 Hz and is linked to active thinking, concentration, and problem-solving. These rhythms are most prominent in the frontal and central areas during wakefulness and alert states.
- Finally, *gamma* waves, which range from 30.0 Hz to 100.0 Hz, represent the fastest EEG rhythms and are associated with advanced cognitive functions like perception, problem-solving, and consciousness. Although these waves are not frequently seen in standard EEG studies, they play a crucial role in specific research settings and are often recorded in different brain regions based on the cognitive tasks being performed.

Once acquired, EEG signals undergo several processing steps. These steps are essential for improving signal quality and extracting relevant information. The processing pipeline ensures that the raw EEG data are transformed to enable monitoring and interpretation of brain activity [16]. The main steps of the EEG signal processing workflow are (Fig. 1.4):

- *Pre-processing:* it is essential to enhance signal quality, especially during online BCI experiments [16]. Temporal and spatial filtering techniques are commonly employed. Temporal filtering typically involves the use of Butterworth or Chebyshev filters in the bands of interest for the type of BCI application, to remove artefacts while preserving relevant information. For spatial filtering, common average referencing and Laplacian filters are frequently used to extract spatial information pertinent to various BCI applications.

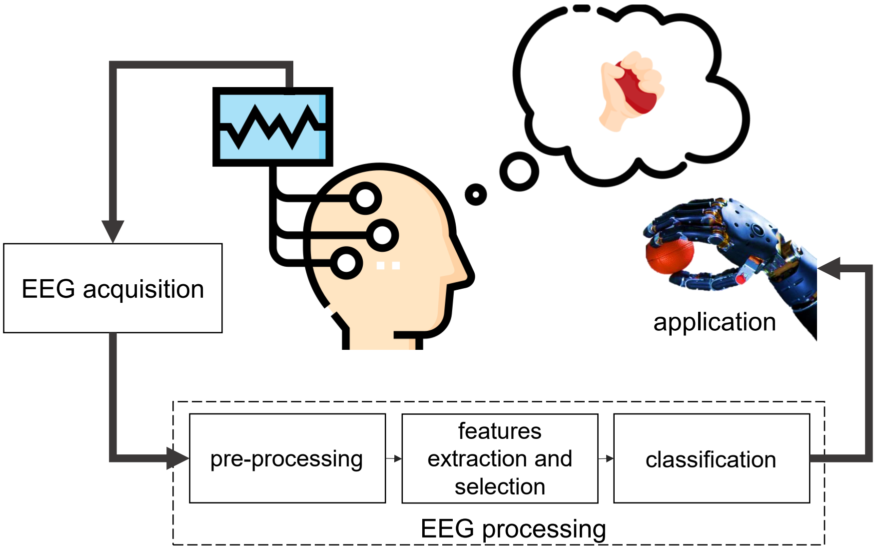


Figure 1.4: Main processing steps of a BCI: from raw EEG signals to control output.

- *Feature extraction:* it is a crucial process in EEG data analysis, as it enables the identification of relevant information within the vast array of data [36]. Various methods are used for feature extraction depending on the specific BCI application. Common approaches include:

  - Power spectral density, frequently used due to its feature stability. It is effective for analyzing stationary signals but not suitable for unstable or time-domain signals [37].
  - Fourier transform is applicable for stationary and narrowband signals, but it struggles with non-stationary signals and is highly sensitive to noise.
  - Common spatial pattern is mainly used for motor imagery-based BCIs. It is sensitive to noise and requires multiple electrodes.
  - Time-domain methods consist in analyzing statistical measures over time such as mean, root mean square, standard deviation, variance, skewness, and kurtosis [16].
  - Time-frequency methods consist in combining temporal and frequency features using short-term Fourier transform and wavelet transforms.

  Often, the extracted features go through a selection process that serves to identify only the most relevant ones by eliminating redundant ones [38].

- *Classification:* it involves categorizing the selected features to interpret brain signals for BCI control. Support vector machines, decision trees, and linear discriminant analysis are commonly used. Deep neural networks such as convolutional neural networks are increasingly being used and can integrate feature extraction, selection, and classification into a single pipeline. These methods, while powerful, risk overfitting due to limited data availability, which can be mitigated by data augmentation and transfer learning strategies.

1.3 Motor imagery paradigm

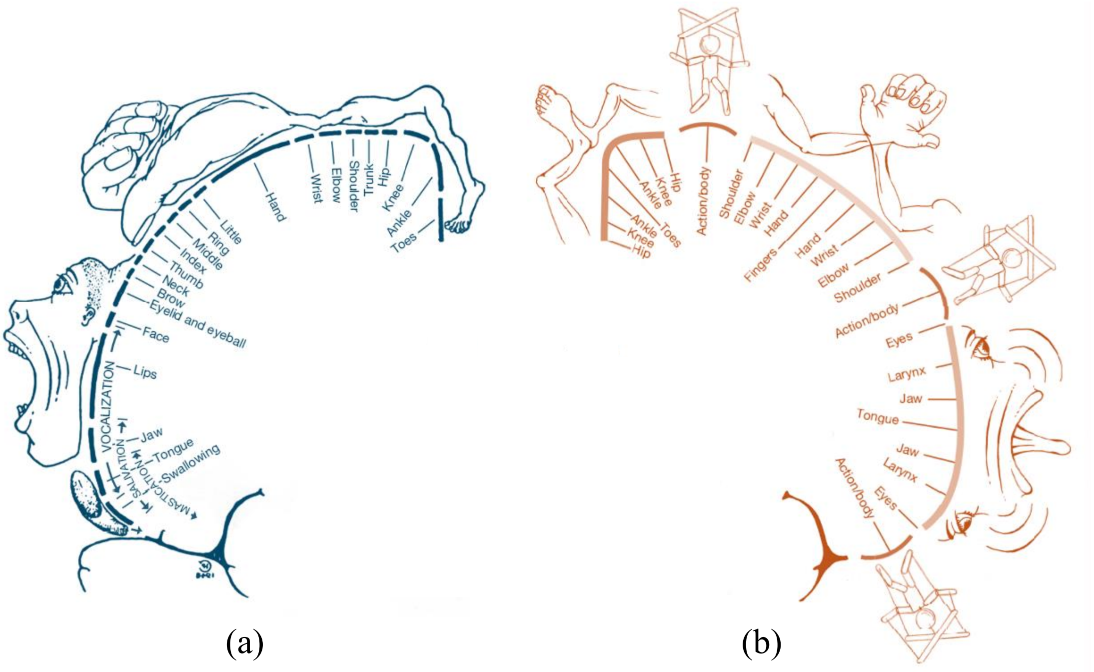
Motor imagery (MI) plays a critical role in BCIs by enabling users to control external devices through mental processes alone. This approach involves mentally practising movements without actually performing them, which creates specific neural patterns that BCIs can recognize [39]. The importance of MI in BCIs spans various fields, including accessibility, rehabilitation, and the integration of advanced technologies.

MI-BCIs rely on spontaneous brain activity linked to mental processes instead of physical actions, engaging the same brain regions as actual movement. This makes them particularly beneficial for individuals with severe physical disabilities [40]. These devices are frequently used in rehabilitation, especially for stroke patients [41]. Recently, there has been a push to apply this technology for the neurorehabilitation of those with neurodegenerative diseases like Parkinson's [42, 43].

In assistive technology, MI-BCIs facilitate the operation of prosthetic limbs, wheelchairs, and computer interfaces, allowing users with physical limitations to carry out daily activities independently [16]. Moreover, MI-BCIs are also finding uses in gaming and virtual reality, where they offer immersive and intuitive control options, thereby enhancing user experience and engagement.

Motor imagery tasks activate specific neural oscillations, particularly within the μ and β frequency bands (from 8 Hz to 30 Hz), which are prominent in the sensorimotor cortex [44]. These oscillations are known as "Event-Related Desynchronizations/Synchronizations (ERD/ERS)" [45]. In particular, when imagining the movement of a specific body part and side, there is a decrease in the μ rhythms of the EEG in the contralateral hemisphere at the beginning of the imagination (a phenomenon known as ERD). Concurrently, there is an increase in the μ band of the EEG under the ipsilateral hemisphere. However, this phenomenon is not visible in all subjects. Ultimately, an increase in β EEG rhythms is evident in the contralateral hemisphere following the conclusion of the imagination of the movement (a phenomenon known as ERS).

To observe these phenomena, it is possible to imagine different types of movements of different parts of the body. MI signals are directly related to the sensorimotor cortex. Historically, the motor cortex organisation was represented by the homunculus model, a continuous map extending from the foot to the face along the precentral gyrus. This representation emerged from early studies where direct cortical stimulation elicited movements primarily from foot, hand, and mouth areas, forming a linear somatotopic map [47]. However, recent studies challenge this traditional view. Using fMRI, a more complex organisation within the motor cortex has been identified [46]. This updated model demonstrates that the classical homunculus is interrupted by regions with distinct connectivity, structures, and functions. These inter-effector regions, located between the representations of the foot, hand, and mouth, feature unique characteristics such as: reduced cortical thickness; strong connectivity with the cingulate-opercular network, which is crucial for action control, physiological regulation, and cognitive functions. This new model suggests that the motor cortex integrates a whole-body action planning system, the somato-cognitive action network, alongside effector-specific regions. Both the model proposed in 1948 and the recently proposed integrated-isolated model are shown in Fig. 1.5 [46]. These models are useful for identifying the points at which electrical activity can be measured to detect specific imaginary movements. For example, electrode Cz can be used to detect imagined foot movements, while electrodes C3, C5, C4 and C6 can be used to detect imagined left and right hand movements.



(a) (b)

Figure 1.5: The old version of the homunculus (a) and the new integrated-isolated model (b) [46].

Despite its potential, motor imagery faces several challenges that need to be addressed to improve its practical applications, particularly in the context of BCIs. The main issues to be addressed are summarised below [48]:

- Developing high-performance MI-BCIs: there is a need for more advanced algorithms capable of accurately discriminating between multiple classes of motor imagery. Enhancing classification accuracy will improve the overall performance of MI-BCI systems.

- Reducing system calibration times: current MI-BCI systems depend significantly on algorithms tailored to individual subjects because of the considerable variability observed between different users and sessions. Implementing transfer learning techniques could substantially reduce or eliminate the lengthy calibration times (typically 15 min to 20 min) required for each user.

- Developing self-paced systems: most existing MI-BCI systems are synchronous, meaning the user is instructed on what to imagine and when. Developing asynchronous systems that operate entirely based on the user's spontaneous intentions would provide a more natural and user-friendly experience.

- Increasing the number of available commands: traditionally, MI-BCI systems have been designed as binary class problems (e.g., imagining left and right hand movements). Expanding the number of classes will allow for more versatile and comprehensive use of this technology.

- Assessing the subject's engagement in motor imagery task: one significant challenge is the lack of direct methods to verify if a subject is performing the motor imagery task as instructed. This inability to objectively confirm compliance can affect the system's reliability.

1.4 The role of metrology

Accurate measurement is the basis for reliable and effective BCI systems, especially when using wireless EEG devices. The characterization of EEG acquisition plays an important role in the development of BCI. Previous studies have demonstrated that consumer-grade EEG devices can be suitable for BCI applications when properly characterised [49]. Specifically, this study found that although these devices exhibit an approximate 10 % gain error, this can be corrected through meticulous calibration. Building on this foundation, this thesis examined the EEG acquisition delay in wireless devices. In real-time applications, it is crucial to take into account not only the mean delay but also the uncertainty associated with it. This consideration is particularly significant for the precise measurement of time-locked neurological events, as uncontrolled variability in delay can undermine the reliability of the analysis.

Signal quality is another critical factor that directly influences BCI performance. EEG signals are highly susceptible to artefacts, which can be classified as physiological (e.g., eye blinks, muscle activity) or non-physiological (e.g., power line interference). Ensuring high-quality EEG data requires effective artefact removal techniques to minimize measurement errors and improve signal integrity. In low-density EEG systems, identifying the minimum number of channels needed for effective artefact rejection is crucial. A key challenge lies in creating techniques that go beyond this threshold, allowing for artefact suppression while still maintaining the system's portability and ease of use. Another challenge is using a quantitative metric to assess how effective artefact removal techniques are. It is essential to determine if these methods can selectively remove artefacts without altering the underlying EEG signal. Despite improvements in signal processing, many studies still rely on visual inspection, highlighting the need for objective and reproducible evaluation criteria.

Another challenge in BCI research is the fragmentation of knowledge across diverse studies, methodologies, and experimental conditions. To drive progress in the field, it is important to adopt a structured approach that preserves, integrates, and consolidates existing findings into knowledge that is both accessible and applicable. Simply reviewing the current state of the art was not enough. A metrological framework allows for the objective comparison of different EEG signal processing strategies, ensuring consistency and reproducibility. This approach facilitates the transition from isolated research findings to reliable and generalizable BCI solutions.

In motor imagery BCIs, it is not trivial to verify whether users actually perform the mental task. Unlike BCIs that depend on external stimuli, where brain responses are distinct and measurable, motor imagery requires individuals to internally generate specific neural patterns. This process is inherently variable and difficult to assess objectively. Users might not fully engage in the task or may lose focus while interacting with the interface, leading to uncertainty in measuring and interpreting EEG signals. By developing robust methods to quantify engagement, BCI paradigms can be improved. This would enhance their practical use, especially in adaptive and real-world applications.

The ultimate goal is to create self-paced BCIs that are robust, accurate, and reliable. However, this remains a major challenge due to the complexities of continuous brain signal processing. It is important to minimize false detections while maintaining system responsiveness. Furthermore, traditional evaluation metrics, such as accuracy, need to be replaced by online evaluation metrics to assess the performance of BCIs in real time.

Chapter 2

Acquisition delay in wireless electroencephalographic instruments

2.1 Rationale for measuring acquisition delay

In recent years, interest in wireless electroencephalographic (EEG) devices has grown due to their increased portability and flexibility. These devices allow for long-term recordings in real-world environments, overcoming the limitations of traditional laboratory-based systems. Wireless EEG devices eliminate the need for physical connections to a computer, allowing for greater mobility [50]. A further advantage of modern wireless EEG devices is the potential for the acquired brain signal to be made available for real-time processing. This represents a significant leap from wired systems, which were primarily designed for accurate signal acquisition without offering real-time processing capabilities. The integration of real-time processing enables feedback, which is a crucial element for closed-loop brain-computer interfaces (BCIs).

However, the transition to wireless technology introduces potential sources of delay in signal transmission. The delay in wired systems was typically negligible [51]. On the other hand, wireless systems inherently introduce delay due to factors such as wireless communication protocols, data buffering, and transmission methods. This would consequently affect the overall performance of a BCI [52, 53]. Acquisition delay can lead to important implications for BCI systems if not properly considered. This is especially true for those that rely on real-time feedback. A notable concern is the impact of delay on the detection and analysis of time-locked neural phenomena. "Event-related potentials" (ERPs) like P300 and "movement-related cortical potentials" (MRCPs) are widely used in BCI applications for their well-defined temporal structure [54]. The P300 is characterized by a high amplitude positive peak that typically occurs around 300 ms following a sensory stimulus. In healthy subjects, the peak may appear within a time window ranging from 300 ms to 600 ms, depending on the placement of the electrodes, the nature of the stimulus, and individual variability [55]. This phenomenon is predominantly observed over midline electrodes. Nevertheless, both the magnitude and latency of this potential tend to increase from frontal to parietal electrode sites [55, 56, 57]. Similarly, MRCP is a low-frequency negative shift in the EEG signal. It emerges in response to motor preparation and execution, with the most prominent changes occurring in the sensorimotor area. MRCP manifest around the onset of movement, with preparatory activity detectable up to 2 s before the action [58]. Recently, MRCPs have gained attention as a key EEG modality for real-time BCI applications, especially in neuromodulation and neurorehabilitation systems, where precise detection is essential for effective outcomes [59, 60].

In applications involving ERPs, averaging over multiple time-aligned trials is a common approach to enhance these components while isolating them from the broader EEG signal. It thus follows that accurate signal timing is essential, as any acquisition delay can disturb the alignment of EEG signals during the mediation process. When an acquisition system introduces delays, it becomes challenging to distinguish between true neural latencies and artificial delays caused by the device. This challenge is heightened in clinical settings, where ERPs, such as P300 and MRCPs, are key in assessing cognitive and motor function or diagnosing conditions such as neurodegenerative diseases. For instance, a delay misinterpreted as neural latency could skew assessments of patient reflexes or cognitive processing speed, leading to potential misinterpretation of data regarding cognitive deficits or motor impairments [56, 61].

Moreover, the assessment of delay is of particular significance for BCI applications when real-time operation is necessary [62]. For example, in communication BCIs for patients with amyotrophic lateral sclerosis, P300-based spellers are commonly used, and any additional delay can significantly impair communication speed and accuracy [63]. Delay fluctuations, also referred to as jitter, can worsen the performance by increasing the variability of ERP components like the P300 peak [64]. Understanding and controlling this delay (and associated uncertainty) introduced by the EEG acquisition is important for ensuring reliable system performance.

Despite the recognized importance of acquisition delay in EEG systems, this aspect remains underexplored in the literature. Most studies focus on the synchronization of stimuli with EEG markers, without thoroughly investigating the delays introduced by the acquisition systems themselves [62, 65]. While some research has proposed customized hardware solutions with controlled delays for specific applications, these are not feasible for most BCI users who rely on commercially available wireless devices [51]. Moreover, there is little work on the uncertainty quantification of these delays, which is essential from a metrological perspective to ensure reliable measurements.

2.2 Measurement station

2.2.1 Design

A measurement station for characterizing the acquisition delay in wireless EEG devices is proposed. The characterization focused on the delay between the acquisition of the EEG signal and its transmission to data processing software. As shown in Fig. 2.1, the proposed measurement station comprises a control unit, a signal generator, and the device under test (DUT). In this setup, the DUT specifically refers to the wireless EEG device under assessment.

The control unit manages the signal generator to deliver a reference voltage waveform to the DUT. The DUT transmits the recorded signal back to the control unit via a wireless connection. Simultaneously, the control unit logs the key events of the reference waveform. The transmission delay can be accurately determined by comparing the signal received via a direct cable connection from the generator with the one transmitted by the DUT. It should be noted that the generator is connected to the DUT through cables. The delay caused by these wired connections is assumed to be negligible compared to the wireless one.

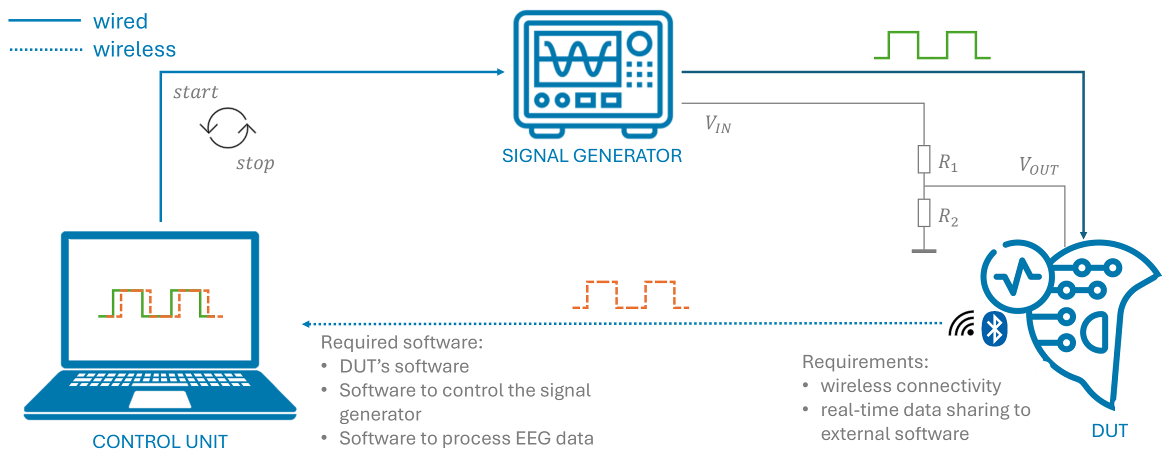


Figure 2.1: Block diagram of the proposed measurement station.

In order to guarantee precise and dependable measurements, it is essential that the proposed measurement station fulfils a number of specific criteria. Firstly, the control unit is required to operate multiple software systems simultaneously. These include a software system for managing the signal generator, a software system for acquiring EEG data in real-time, and, where applicable, the proprietary software of the DUT. It is worth noting that the software used to control the signal generator and acquire EEG data may be integrated into a single software system. Furthermore, the temporal interval between consecutive key events in the reference waveform has to be sufficiently longer than the maximum expected acquisition delay. This requires an initial estimation of the delay.

In order to be included in the analysis, EEG devices have to be equipped with wireless connectivity and support real-time data sharing outside their proprietary software. Furthermore, the signal generator is required to produce voltage waveforms within the typical amplitude range of EEG signals, which is approximately 0.5 µV to 100.0 µV [32]. This is to prevent device saturation or potential damage. If necessary, a voltage divider can be used to adjust the amplitude of the test signals.

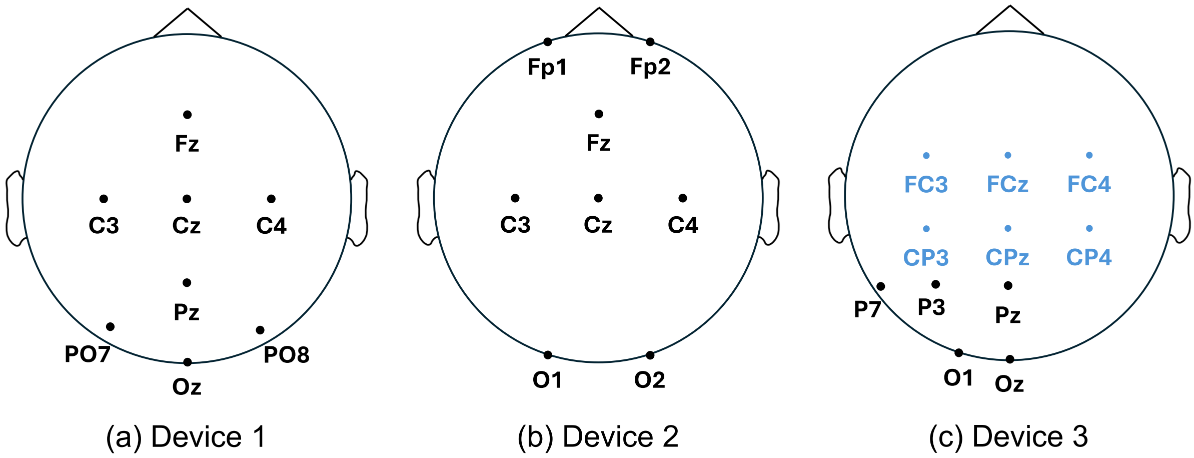
2.2.2 Implementation

The proposed measurement station was employed for the assessment of the performance of three wireless EEG devices, each differing in configuration and capabilities. Device 1 features eight monopolar EEG channels and operates at a fixed sampling rate of 250 Hz. Device 2, with eight monopolar EEG channels, offers a configurable sampling rate of either 512 Hz or 1024 Hz. Device 3 combines five monopolar and three bipolar EEG channels, operating at a fixed sampling rate of 125 Hz. Simulink (by MathWorks) was selected as the software for data acquisition and processing due to its widespread use in brain-computer interface applications. Devices 1 and 3 are equipped with application programming interfaces (APIs) that support direct integration with Simulink for real-time data acquisition. Conversely, Device 2 lacks a Simulink-compatible API, requiring the use of its proprietary software for signal acquisition. The EEG data can then be transferred to Simulink via the Lab Streaming Layer (LSL) protocol. The main features of the three EEG instruments are summarized in Tab. 2.1, and their electrode placements are illustrated in Fig.2.2.

The Agilent 33220A was employed as signal generator. The generator operates within

Table 2.1: Overview of the included EEG instruments

	Channels	**Sampling rate** (Hz)	**API availability**	**Connection to Simulink**
Device 1	8 monopolar	250	Yes	Direct via API
Device 2	8 monopolar	512 or 1024	No	Via LSL
Device 3	5 monopolar, 3 bipolar	125	Yes	Direct via API



(a) Device 1 (b) Device 2 (c) Device 3

Figure 2.2: Electrode placement of the three EEG instruments: monopolar electrodes are shown in black, and bipolar electrodes in blue.

an amplitude range of $20\,\mathrm{mV_{pp}}$ to $20\,\mathrm{V_{pp}}$ in High-Z mode. Nevertheless, EEG signals typically exhibit smaller amplitudes. Thus, a voltage divider was introduced to scale the output to appropriate levels, configured as follows:

- Input voltage: $500\,\mathrm{mV_{pp}}$
- $R_1 = 1\,\mathrm{M\Omega}$
- $R_2 = 100\,\Omega$

The resulting partitioned voltage was measured at approximately $47.5 \pm 0.1\,\mu\mathrm{V}$, ensuring compatibility with EEG signal characteristics. A square wave was selected as the reference waveform due to its sharp transitions, which served as precise markers for latency measurement. It is important to note that the rise and fall times of the Agilent generator are on the order of nanoseconds, effectively rendering these transitions instantaneous when measuring delays with millisecond-level accuracy.

The signal generator was controlled via a custom-designed LabVIEW model. Each cycle entailed switching the generator on, acquiring signals with the DUT for a defined duration, and then switching the generator off. Based on preliminary tests, the worst-case acquisition delay was estimated to be approximately 200 ms. Consequently, the signal generation window was set at 120 s between consecutive on-off commands, a duration that greatly exceeded the anticipated delay to ensure robust measurements.

To evaluate the mean transmission delay and estimate its uncertainty, multiple measurements were performed. The DUT was operated continuously from a fully charged battery until depletion, with the signal generator repeatedly activated and deactivated.

The pertinent events of the reference waveform were defined as the time points at which the generator was commanded to be switched on or off. Furthermore, the status of the generator was queried via Simulink using the User Datagram Protocol (UDP) to confirm the correct execution of the command, thus ensuring the accurate recording of the edges of the reference waveform.

To further ensure the reliability of the LabVIEW-Simulink integration, repeated ping-pong tests were conducted to assess communication delays. These tests revealed a half round-trip time of $30 \pm 2\,\mu\text{s}$. As a result, it was concluded that this delay was negligible and did not affect the measurement system's accuracy when using the same PC for both LabVIEW and Simulink.

The developed Simulink model comprised two primary components:

- A UDP read block for receiving the signal generator's state from LabVIEW.
- A block for acquiring the EEG signal from the DUT.

For Devices 1 and 3, the EEG data was received via the provided APIs, while for Device 2, a custom Level-2 MATLAB S-function was implemented to enable data acquisition. Although an open-source Simulink library for LSL signal acquisition (SimBSI) was available, it was not used due to observed stability issues [66]. The acquired signals from both blocks were sent simultaneously to a scope for real-time monitoring and stored for subsequent offline processing. To ensure synchronisation, the Simulink model was initiated prior to launching the LabVIEW model. This systematic approach enabled accurate characterization of the acquisition delay between the EEG instrument and the data processing software, providing reliable metrics for evaluating wireless EEG systems' performance.

2.2.3 Assessment of acquisition delay

The acquisition delay and its associated uncertainty for wireless EEG instruments were determined through offline analysis using MATLAB©. For each DUT, the acquired data was stored in a file containing two primary datasets: the on/off status of the signal generator at each time sample (with values of +1 and -1, respectively) and the signal acquired by the DUT over the same temporal sequence.

In analysing the signal generator's status, transitions from -1 to +1 were designated as “start” timestamps, while those from +1 to -1 were marked as “stop” timestamps. The aforementioned timestamps corresponded to the generator switching on and off, respectively. Meanwhile, the signal acquired by the DUT exhibited a periodic alternation between 120 s of baseline activity (generator off) and 120 s of square waves (generator on). Fig. 2.3 shows an example of the difference between the generator step (in orange) and the square waves acquired by the DUT (in blue).

In order to identify transitions in the acquired signal, thresholds were applied in order to distinguish between baseline and square wave segments. The baseline segment was primarily characterized by noise, with its amplitude consistently observed between the 45^{th} and 55^{th} percentiles of the acquired signal. This range encompasses the most frequently occurring signal values. In contrast, the highest and lowest values of the square wave segment were observed below the 5^{th} percentile and above the 95^{th} percentile, respectively. Based on this distribution, positive transitions (baseline to square wave) were defined as instances where the signal exceeded the 75^{th} percentile, which lies midway

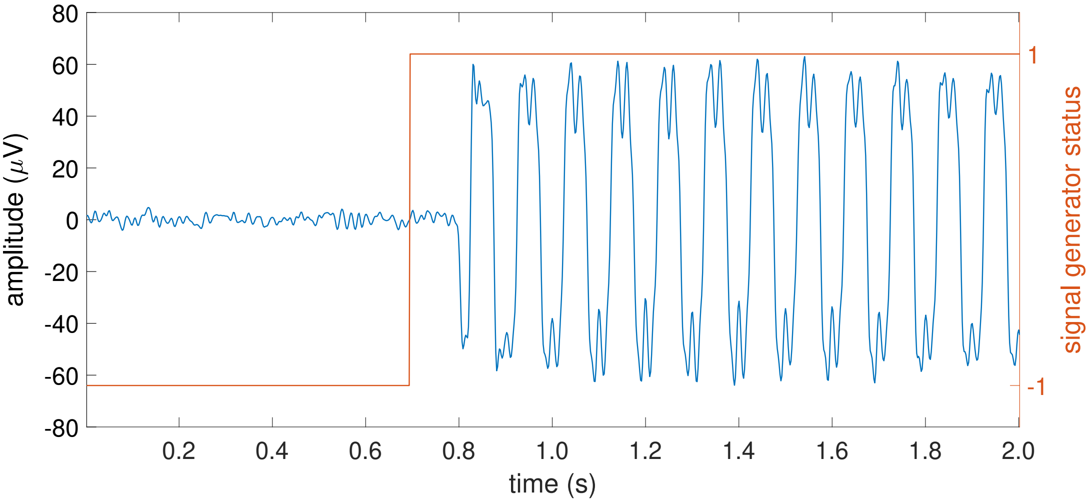


Figure 2.3: Example of the difference between the generator step (in orange) and the square waves acquired by the DUT (in blue).

between the 55th and 95th percentiles. Similarly, negative transitions (square wave to baseline) were identified when the signal dropped below the 25th percentile, which is midway between the 45th and 5th percentiles.

The acquisition delay was ultimately calculated by comparing the time samples of the detected transitions in the signal generator's output with those in the DUT's acquired signal.

2.2.4 Results

The measured acquisition delays for the three wireless EEG instruments are summarized in Fig.2.4. These measurements were taken at both the start and stop transitions of the signal generator, providing a comprehensive view of the delay distributions for each instrument. Each box plot represents the range of repeated delay measurements, emphasizing variability and central tendencies.

For Device 1, the results indicate the lowest mean delay, approximately 20 ms. At the signal generator's start, the delay corresponded to around six samples, while at the stop it reduced slightly to about five samples. These values confirm Device 1's suitability for real-time EEG applications, as its minimal delay ensures compatibility with time-sensitive processing tasks.

Device 2, however, demonstrated significantly higher delays, particularly when operating at a sampling rate of 512 Hz. Under these conditions, the mean delay reached 102 ms, equivalent to approximately 54 samples, with slightly more delay observed at the start than at the stop. Increasing the sampling rate to 1024 Hz reduced the mean delay to 88 ms, showing some improvement. Nevertheless, Device 2 consistently exhibited the poorest performance among the instruments. A key factor contributing to its delay is the absence of a direct API for communication with the control unit. Instead, EEG data must first pass through proprietary software before being forwarded to the acquisition and processing environment (Simulink in this study), which introduces an additional layer of latency.

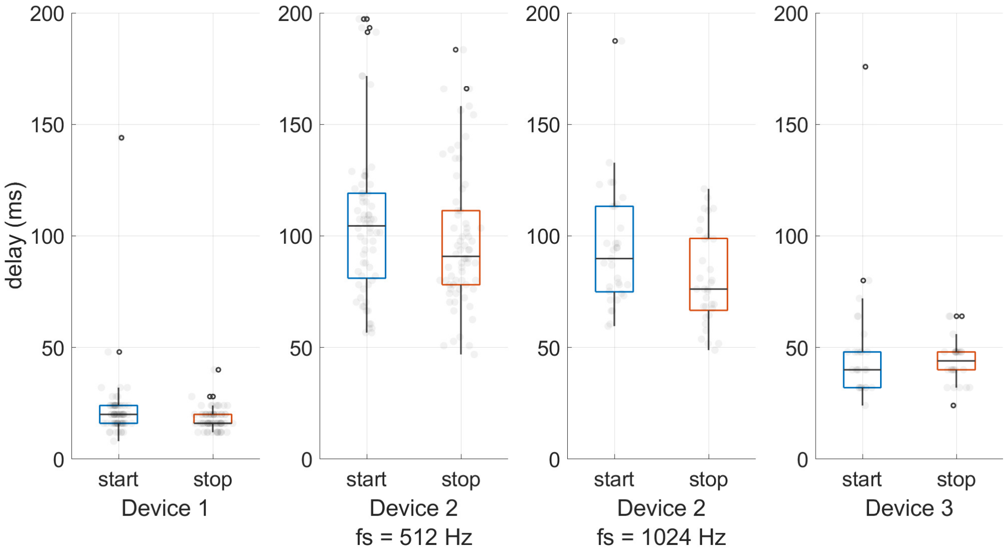


Figure 2.4: Acquisition delays in the processing software for each of the three devices under test. The delays were measured at both the start and the stop of the signal generator.

Device 3 exhibited intermediate performance, with a mean delay of 47 ms. Its performance remains acceptable for real-time applications and its behavior aligns closely with that of Device 1.

The delays measured at the start and stop of each generated wave were found to be consistent across all instruments, thereby indicating the absence of data loss. This conclusion was corroborated by an analysis of the half-period of the square waves captured by the EEG instruments, which exhibited a consistent measurement of 50 ms across all instruments. Moreover, the mean duration of the start-stop cycles was found to be 120 s, confirming the absence of packet loss during testing.

The measured delays, half-period durations of the square waves, and the total start-stop cycle durations for the three wireless EEG instruments under test are summarized as follows:

- **Device 1:**
  - Delay at start: 22 ± 2 ms
  - Delay at stop: 18 ± 1 ms
  - Half-period of the square waves: 50 ± 4 ms
  - Start-stop cycle duration: 120.09 ± 0.02 s
- **Device 2** (512 Hz)**:**
  - Delay at start: 106 ± 4 ms
  - Delay at stop: 98 ± 4 ms
  - Half-period of the square waves: 50 ± 2 ms
  - Start-stop cycle duration: 120.03 ± 0.02 s
- **Device 2** (1024 Hz)**:**

  - Delay at start: 94 ± 5 ms
  - Delay at stop: 81 ± 4 ms
  - Half-period of the square waves: 50 ± 1 ms
  - Start-stop cycle duration: 120.03 ± 0.02 s

- **Device 3:**
  - Delay at start: 50 ± 6 ms
  - Delay at stop: 43 ± 2 ms
  - Half-period of the square waves: 50 ± 4 ms
  - Start-stop cycle duration: 120.07 ± 0.03 s

In order to demonstrate the statistical significance of the aforementioned differences, the "analysis of variance" (ANOVA) was employed for data that exhibited a normal distribution, whereas the Kruskal-Wallis test was employed for data that did not (Tab. 2.2). The normality of the data was verified using the Shapiro-Wilk test, with a significance level of $\alpha = 5\,\%$.

Table 2.2: Statistical comparison of start and stop delays between different EEG instruments. ANOVA was applied for normal distributions, otherwise the Kruskal-Wallis test. The test significance was fixed at $\alpha = 5\,\%$. The resulting p-values are reported.

Devices		Start	Stop
Device 1	Device 2, fs = 512 Hz	< 0.01	< 0.01
Device 1	Device 2, fs = 1024 Hz	< 0.01	< 0.01
Device 1	Device 3	< 0.01	< 0.01
Device 2, fs = 512 Hz	Device 2, fs = 1024 Hz	FAIL	FAIL
Device 2, fs = 512 Hz	Device 3	< 0.01	< 0.01
Device 2, fs = 1024 Hz	Device 3	< 0.05	FAIL

The results highlighted significant differences in delays between: Device 1 and Device 2 (for both sampling rates), Device 1 and Device 3, Device 2 at 512 Hz and Device 3. Interestingly, the tests failed to reject the hypothesis of no significant differences when comparing the two sampling rates of Device 2, supporting the hypothesis that its primary delay source arises from the lack of direct communication with the control unit. However, at 1024 Hz, Device 2 demonstrated some compatibility with the delays of Device 3, particularly during stop transitions.

Additionally, paired tests were conducted to compare start and stop delays within each instrument (Tab. 2.3). Depending on data normality, either paired t-tests (for normal distributions) or Wilcoxon signed-rank tests (for non-normal distributions) were used. These tests revealed a statistically significant difference in start and stop delays only for Device 1 ($\alpha = 5$ %), while no significant differences were observed for Device 2 or Device 3.

A noteworthy observation is the consistently higher delay measured at the first start of the signal generator in all instruments. In particular, the delay between the activation of the signal generator and the detection of the first square wave by the EEG instrument

Table 2.3: Statistical comparison of paired start and stop delays for the same instrument. A paired t-test was used for normally distributed data, while the Wilcoxon signed-rank test was applied for non-normal distributions. The test significance was fixed at $\alpha = 5\,\%$. The table reports whether the start delay was less than, equal to, or greater than the stop delay, along with the associated p-values.

Devices	Comparisons	p-value
Device 1	start > stop	0.02
Device 2, fs = 512 Hz	start = stop	0.25
Device 2, fs = 1024 Hz	start = stop	0.09
Device 3	start = stop	0.74

was greater during this initial event than during subsequent starts of the signal generator. This anomaly was traced back to the initialization phase of Simulink. Once the system stabilized, the delays became shorter and more consistent, further supporting the robustness of the measurement system during steady-state operation. Fig. 2.5 illustrates this behavior for Device 1, where the delay measured at the first start reached approximately 144 ms before stabilizing to significantly lower values for subsequent activations of the signal generator.

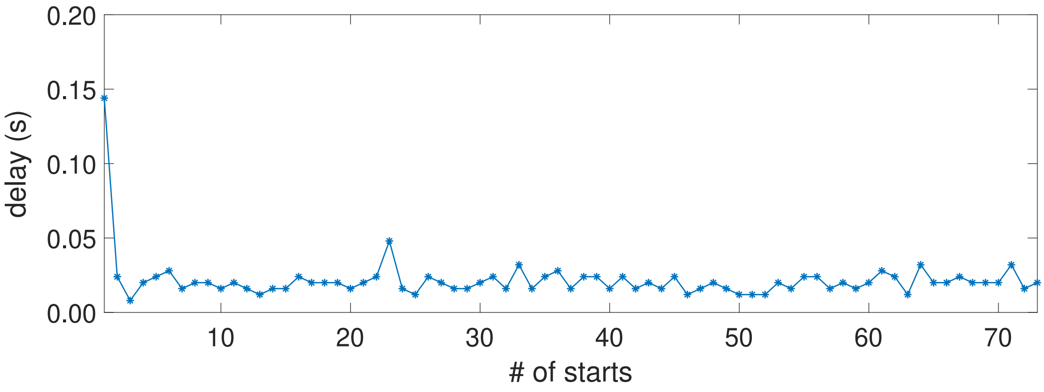


Figure 2.5: Example of measured delays at signal generator starts for Device 1, highlighting the higher delay at the first activation.

2.3 Impact on time-locked phenomena

Time-locked neural phenomena are crucial indicators in the analysis of EEG signals, as they arise in response to specific stimuli or events. This section presents a BCI experiment that illustrates how delays in signal acquisition can impact the measurement of these phenomena, potentially undermining the effectiveness of BCI systems. Key phenomena such as the P300 and the MRCP are particularly significant in BCI applications. The three DUTs examined in the previous analysis were investigated here under controlled experimental conditions, specifically in a BCI focused on detecting left and right hand movements.

2.3.1 On-field experiment

The BCI system used to assess the impact of the acquisition delay consists mainly of two hardware components: the DUT, which will refer to here as the EEG acquisition instrument, and the control unit, which will refer to here as the personal computer. The EEG instrument transmits data to the personal computer via Bluetooth. The personal computer handles the acquisition and processing of the EEG data as well as the timing of the paradigms used. Simulink was chosen as the software for EEG data acquisition and processing. Unity[1] was chosen to manage the timing of the BCI paradigm. If necessary, the proprietary EEG instrument software was also run in parallel with Simulink on the personal computer. Communication between Simulink and Unity was implemented via the UDP protocol. The transmission delay via UDP was considered negligible compared to the acquisition delay as both software were run on the same personal computer.

The BCI system was designed to detect neural phenomena associated with right- and left-hand movements, focusing on the P300 and MRCP components. A synchronous paradigm was used to highlight the effects of acquisition delays. In this setup, the timing of events was predetermined and controlled externally, ensuring accurate synchronization between the EEG signals and the stimulus cues. Each trial followed these steps (Fig. 2.6):

- Fixation cross: A cross shown for 2.00 s.
- Visual cue: An arrow indicating left or right, displayed for 1.25 s. Specifically, when the arrow pointed to the right, participants were instructed to move their right hand and vice versa.
- “GO!” signal: Shown from 3.00 s to 6.00 s, prompting participants to perform the movement.
- “RELAX” signal: Displayed for a random duration between 1.00 s to 2.00 s to avoid bias in consecutive trials.

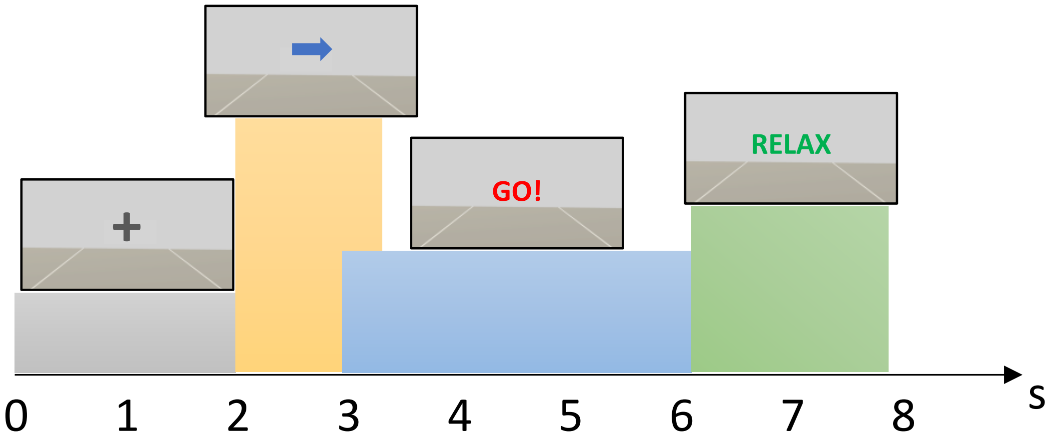


Figure 2.6: Timing of a single trial.

Participants were instructed to get ready for the movement when the visual cue appeared and to carry out the movement upon seeing “GO!”. Each experimental session

---

[1] https://unity.com/

included three runs, with each run containing 30 randomized trials (15 trials for each hand). EEG data and timing information from the paradigm were recorded for offline analysis.

2.3.2 Analysis

The offline analysis of EEG signals was performed in MATLAB© to evaluate the effects of the acquisition delay measured in the previous section. Each analysis concentrated on particular time-locked events and relevant acquisition channels. In exploring the neural phenomena present in the EEG signals, instrument-specific measured delays were included in the analysis to evaluate their impact on the detectability of these neural phenomena. The goal was to demonstrate that these phenomena become more evident when the delay can be at least partially corrected. Conversely, this implies that the detection of neural phenomena is hindered by delays that can not be corrected.

The P300 component was examined in relation to the visual cue (2 s), which served as a "go/no-go" stimulus. This visual cue was considered a significant event as it prompted the participant to get ready for the next action. While several visual stimuli were presented during each trial, only the visual cue elicited the preparatory neural response. The analysis focused on EEG channels in the parieto-occipital region, which is particularly important for identifying the P300 component. For each EEG instrument, the channels in this area were chosen. However, due to the fixed configurations of different instruments, it was not always feasible to obtain the same set of channels across all instruments. The data processing steps were as follows:

- filtering EEG signals in the range from 0.5 Hz to 15.0 Hz using a 4th-order zero-phase Butterworth filter [67];
- segmenting the data into epochs from 200 ms before to 1000 ms after the visual cue;
- averaging all epochs to enhance the signal-to-noise ratio.

The peak amplitude of the P300 was identified as the highest deflection observed within the selected time window. The latency of the P300 was measured as the time taken to reach the peak in relation to the onset of the visual cue. Finally, this latency was corrected for the measured delay of each EEG instrument.

The MRCPs were examined in relation to the "GO!" cue (3 s). Signals were gathered from EEG channels located in the sensorimotor area for each instrument, which plays a key role in motor preparation and execution. The processing steps included [68]:

- initial filtering of EEG signals within the 0.3 Hz to 70.0 Hz range was performed using a 4th-order zero-phase Butterworth filter;
- artifacts were identified and marked with EEGLAB [69], based on criteria such as signals exceeding ± 200 µV or showing joint probability or kurtosis that were more than five standard deviations from the mean;
- the final filtering of signals was conducted in the 0.5 Hz to 5.0 Hz range to isolate the low-frequency components typical of MRCPs [70];
- epochs were segmented from 500 ms before to 500 ms after the "GO!" cue to capture the negative peak linked to movement onset [58, 68, 70];

- epochs marked with artifacts were discarded, and the remaining epochs were averaged.

For MRCP analysis, the peak amplitude was identified as the lowest deflection within the selected time window. The latency of the MRCP was determined as the time of the peak relative to the "GO!" cue, and this latency was adjusted based on the measured delay of each EEG instrument.

2.3.3 Results

A 31-year-old male participated in the experiments. Before taking part in the study, the participant was required to sign an informed consent form approved by the Ethical Committee of the University of Naples Federico II (Ethical Committee of Psychological Research, protocol number 24/2024). To reduce potential bias, all EEG instruments were used on the same day, and the order of their use was randomized. Each session lasted about 30 minutes per instrument.

During the experiments, the participant was comfortably seated in a chair with both arms extended on a table. The participant was instructed to perform repeated hand-squeezing movements with either the right or left hand, while holding a soft ball in each hand. This motor task was selected to ensure consistent muscle activation and reliable detection of motor-related EEG signals. To illustrate how acquisition delay affects the detection of EEG phenomena, two instruments were chosen based on their performance: Device 1 and Device 2, with the latter operating at a sampling rate of 512 Hz. These instruments were identified as the best and worst in terms of acquisition delay from the analysis conducted in the previous section.

As shown in Fig. 2.2, the two instruments have EEG channels located in different positions. In order to determine the P300 latency under similar conditions, the focus was on the occipital and parieto-occipital regions. Specifically, PO7, Oz, and PO8 were chosen for Device 1, and O1 and O2 for Device 2. This choice aims to minimize variability caused by differences in electrode placement and signal propagation across the scalp [56].

Fig. 2.7 shows the 90 trials in a color-scale plot acquired with Device 1 (top) and the mean signal of all trials along with the standard deviation of the mean (bottom) per each EEG channel. The time window analyzed ranges from 200 ms before the visual stimulus (cue time) to 1 s after the stimulus. While the primary focus is on the parieto-occipital channels, the figure incorporates all available channels in accordance with the standard 10-20 electrode placement system. This comprehensive representation ensures that the observed phenomenon in the area of interest is not incorrectly attributed to other processes, such as motor-related activity. As illustrated in the figure, channels outside the parieto-occipital region do not show the characteristic peak around 300 ms after the stimulus. In each plot, the dashed vertical black line represents the reference point for the onset of the visual stimulus (cue), whereas the dashed vertical red line shows the latency of 300 ms following the stimulus.

Fig. 2.8 presents the same information for Device 2.

Both the aligned trials and the averaged signal show a positive peak amplitude occurring around 300 ms after the cue. Notably, Device 1 has better alignment of the peaks across individual trials compared to Device 2. This enhanced alignment is likely due to the lower uncertainty in delay measurement seen with Device 1 when compared to Device 2.

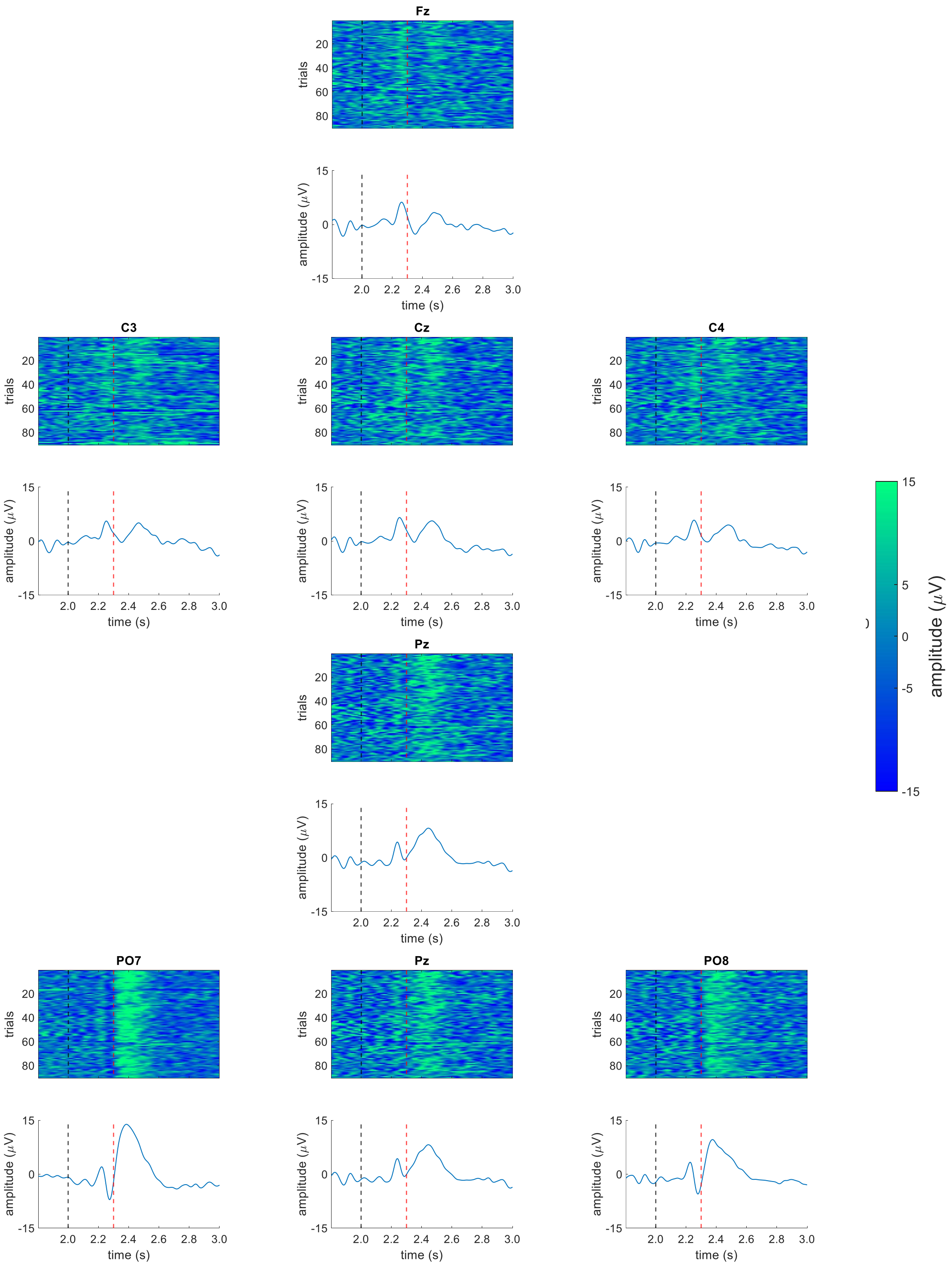


Figure 2.7: Results of the P300 analysis for Device 1. For each EEG channel, the upper plot shows the 90 filtered epochs as color-scale graphs, while the lower plot shows their corresponding mean and the standard deviation of the mean associated with them. The dashed black line indicates the cue event, and the red line marks the expected P300 latency, which occurs 300 ms after the cue.

Additionally, while the P300 peak was identified with both instruments, the latency for Device 2 was significantly longer than that for Device 1. To explore the reason behind this, the observed latencies were adjusted by subtracting the respective measured delay for each instrument. The mean delay values reported in Sec. 2.2 were used for this adjustment. The corrected latency values, shown in Table 2.4. These corrections show that the latency values for the P300 peak, after accounting for instrument-specific delays, are compatible between the two instruments, with only minor deviations of a few milliseconds. This indicates that the differences in raw latencies were mainly a result of variations in instrument delay rather than being inherent to the EEG signal itself.

Table 2.4: P300 peak latencies for each instrument before and after applying delay correction. The focus is on parieto-occipital channels.

Device	Channel	Peak Latency (ms)	Corrected Latency (ms)
Device 1	PO7	384	364
	Oz	376	356
	PO8	376	356
Device 2	O1	462	360
	O2	454	352

In the second part of the analysis, the focus shifted to detecting MRCPs in the sensorimotor area. The EEG channels of interest were C3, Cz, and C4, which correspond to the left motor cortex, the midline, and the right motor cortex, respectively. Notably, these three channels were available for both instruments, enabling a direct comparison under the same conditions. Unlike the P300, MRCPs have a weaker amplitude and are more susceptible to the mean acquisition delay and its associated uncertainty. Therefore, only the sensorimotor channels were analysed, as examining other regions would not provide relevant insights into the observed phenomenon.

Fig. 2.9 shows the results of the MRCP analysis for Device 1 (Fig. 2.9a) and Device 2 (Fig. 2.9b). For each EEG channel, the upper section features a color-scaled plot representing the 90 aligned and bandpass-filtered epochs, while the lower section illustrates the grand average of the MRCPs and its standard deviation of the mean. The time window extends from 0.5 s prior to the nominal movement onset ("GO!" signal) to 0.5 s following it. The nominal movement onset is indicated by the black dashed vertical line at 3.00 s.

For Device 1 (Fig.2.9a), a clear negative peak is observed at the movement onset, particularly evident in the Cz channel, with consistent patterns also present in C3 and C4. This demonstrates that Device 1 can reliably capture MRCPs despite the inherently low amplitude of this signal. In contrast, the data acquired with Device 2 (Fig.2.9b) does not exhibit a distinct MRCP. The lack of consistent patterns across C3, Cz, and C4 highlights the influence of the higher acquisition delay and its associated uncertainty on the detection of this phenomenon.

2.4 Discussion

The findings emphasize the importance of accurately characterizing acquisition delay in wireless EEG instruments, particularly when performing time-sensitive analyses of EEG phenomena. There were notable differences among the devices under test, with Device 1 showing a lower and more consistent delay than the others. Notably, the instruments were treated as "black boxes", meaning there was no prior knowledge of their hardware setups or transmission methods.

The challenge of acquisition delay in wireless EEG systems is well-recognized in existing literature, often tackled through hardware adjustments [51], synchronization techniques for multiple instruments [71, 72], or the implementation of external triggers [62,

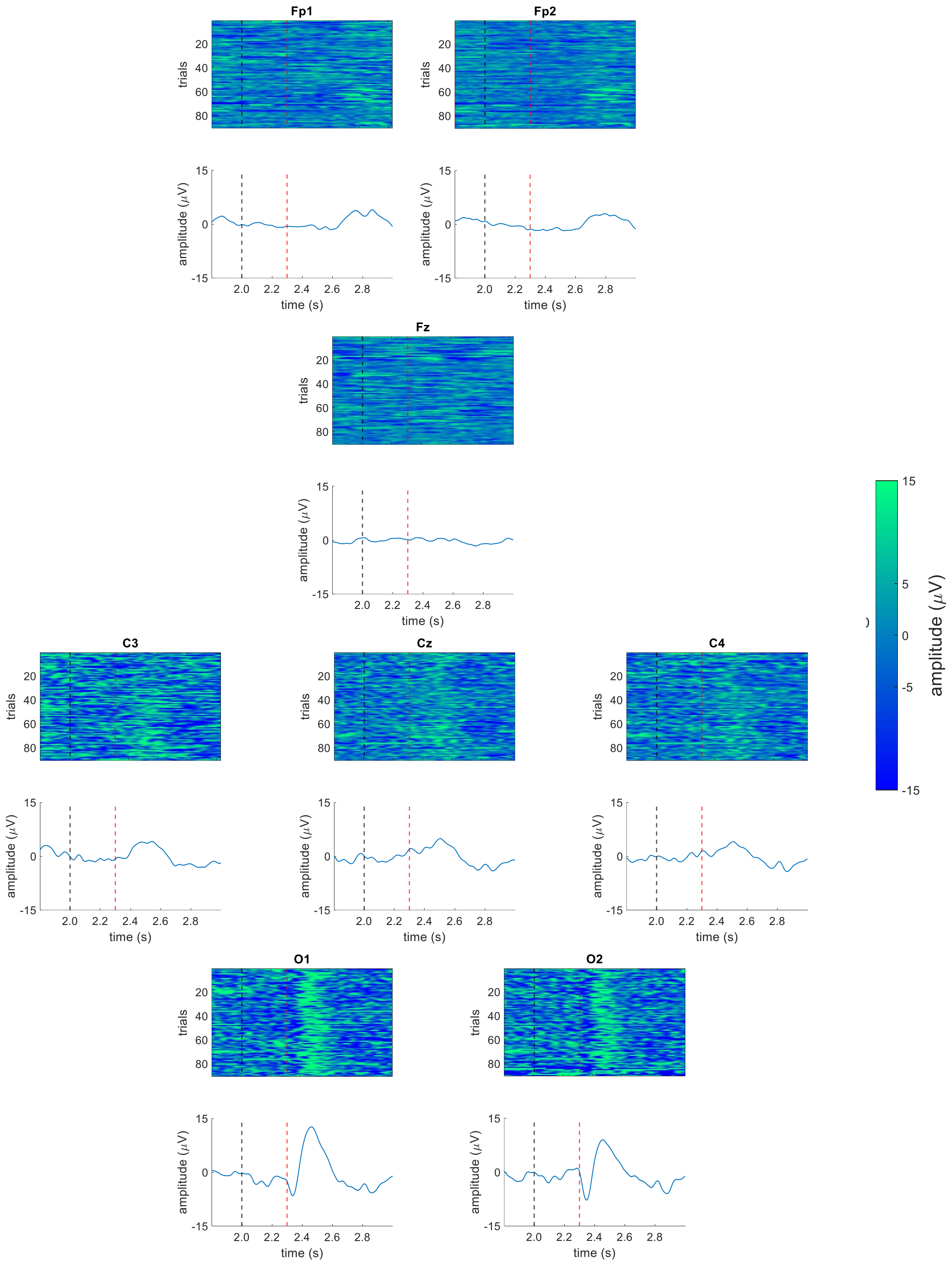
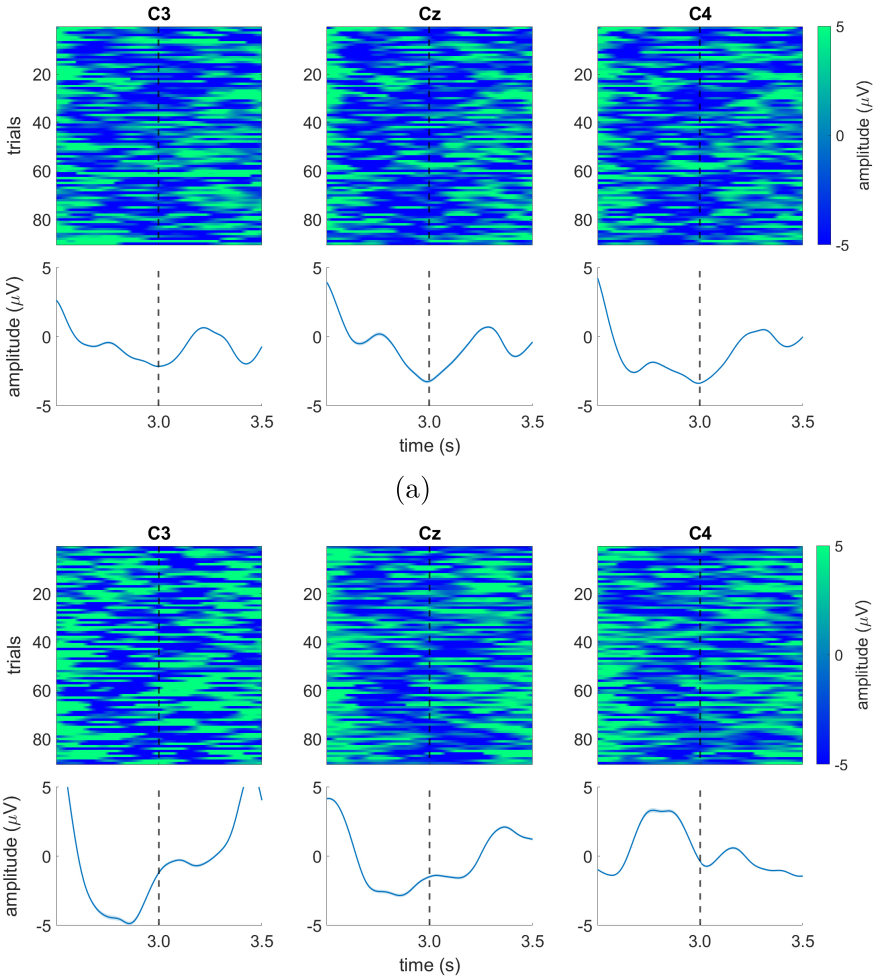


Figure 2.8: Results of the P300 analysis for Device 2. For each EEG channel, the upper plot shows the 90 filtered epochs as color-scale graphs, while the lower plot shows their corresponding mean and the standard deviation of the mean associated with them. The dashed black line indicates the cue event, and the red line marks the expected P300 latency, which occurs 300 ms after the cue.

65]. While these approaches can help reduce delay variability, they cannot completely eliminate the inherent uncertainty that comes with real-time data acquisition. Therefore, a comprehensive characterization of delay is particularly important for consumer-grade wireless EEG instruments, as such variability can greatly affect the reliability of the data.



(a)

(b)

Figure 2.9: Results of the MRCP analysis focusing on the sensorimotor area channels C3, Cz, and C4: (a) Device 1, (b) Device 2. For each EEG channel, the upper plot shows the 90 filtered epochs as color-scale graphs, while the lower plot shows their corresponding mean and the standard deviation of the mean associated with them. The dashed black line indicates the “GO!” signal at 3.00 s.

This work highlights that after measuring the mean delay, it is essential to correct it during post-processing to improve data accuracy. However, the uncertainty associated with delay, which appears as noise, can not be completely removed. This presents a challenge in identifying neural phenomena like time-locked ERPs and MRCPs. The higher the delay uncertainty, the more difficult it is to accurately detect these phenomena. Consequently, delay variability becomes a crucial factor in assessing instrument performance.

The effect of delay on P300 detection is especially significant given the common use of this ERP in cognitive and clinical assessments. In healthy individuals, the P300 latency generally ranges from 300 ms to 600 ms [55], with variations influenced by factors such as age and task complexity. However, the delay of about 100 ms measured for Device 2 could lead to incorrect conclusions, potentially causing misdiagnoses of cognitive impairments in individuals who are functioning normally [73]. By implementing delay corrections, even instruments with considerable inherent delays could be suitable for P300 analysis.

Additionally, slight differences in P300 latency and amplitude between instruments might also be affected by variations in the participants' conditions during the experiments, as these were conducted consecutively rather than simultaneously [72].

In the case of MRCP detection, the differences between the instruments were more noticeable. Even though the same experimental setup, participant, and signal processing pipeline were used, only Device 1 consistently captured the negative deflection characteristic of MRCPs. In contrast, Device 2 did not detect any clear MRCP patterns, likely due to its higher delay uncertainty and the inherently low amplitude of MRCPs, which makes them especially vulnerable to timing inaccuracies. This limitation highlights the essential role of acquisition delay characterization in ensuring the reliability of EEG-based studies.

Many MRCP studies typically align trials with an external trigger that marks the actual start of movement, rather than depending solely on the nominal onset. This synchronization, which can be achieved using exoskeletons, motion sensors, or cameras, has been shown to improve MRCP detection [68, 70]. However, in this study, no external reference was utilized, which allowed for slight variations in the actual movement onset to influence MRCP detection. These variations were particularly noticeable with Device 2, while Device 1 still managed to capture MRCPs. By not relying on external synchronization, this study underscores the additional challenges that delay uncertainty brings when processing low-amplitude neural phenomena like MRCPs.

Moreover, the protocol used in this study separated the cue and "GO!" signals to prevent overlap between P300 and MRCP components. This separation allowed for a clearer distinction between cognitive and motor-related potentials. Future research could investigate increasing the temporal gap between these signals or incorporating external markers to indicate the actual onset of movement for enhanced precision.

The choice to focus on executed rather than imagined movements was intentional. BCIs based on motor imagery are widely used. However, they involve an additional source of uncertainty because there is no clear time marker indicating when and if the user actually starts the mental task. From a measurement point of view, this represents a significant limitation, as it makes it difficult to align EEG signals with the intended event. On the other hand, the performed movements provide a clear temporal reference that reduces ambiguity and allows for a more accurate assessment of delay uncertainty.

This study demonstrates that by combining delay characterisation, appropriate post-processing corrections and well-structured experimental protocols, the challenges associated with acquisition delays in wireless EEG instruments can be greatly reduced. By addressing these issues, the reliability of EEG-based neural search can be improved.

Chapter 3

Artefact removal for low-density electroencephalography

The widespread integration of Brain-Computer Interface (BCI) applications into daily life remains limited. This is primarily due to the need of enhancing electroencephalographic (EEG) signal measurements, particularly in the context of low-density EEG. Within this realm, brain signals are mixed with extraneous noise. The identification and characterisation of these interference signals are of crucial importance, serving as a pivotal pre-requisite for the accurate interpretation and cleaning of EEG data.

To this end, an analysis of current artefact removal techniques was carried out. The effectiveness of these methods and their limitations were examined. Furthermore, a novel approach was proposed to address the challenges associated with low-density artefact removal.

Part of the work described in this chapter has been previously published in [74, 75].

3.1 Artefacts

Out of a research laboratory, EEG signals often encounter significant contamination from various sources. The EEG electrodes acquire signals from both non-physiological and physiological sources, each leaving a distinct imprint on the recorded signal.

Non-physiological sources of noise originate from both environmental and experimental sources [76, 77, 78]. Environmental artefacts often stem from the surroundings and can pose challenges in EEG recordings. A common environmental artefact is power supply interference, typically occurring at 50-60 Hz due to the electrical wiring system. To mitigate this, a notch filter can be applied at the specific frequency. Electromagnetic interference, radio frequency interference and parasitic capacitive couplings, arising from various sources such as instrumentation, circuit components, and room wiring, also contribute to the complexity of the recording process.

Strategies to deal with environmental interference include the physical arrangement of cables, shortening and twisting them, and the use of electromagnetic shielding.

Incorrect procedural setup (e.g., improperly applied electrodes) can yield inconsistent signals. Subject motion poses another challenge, leading to uncontrolled changes in contact impedance at the electrode-skin interface and causing signal distortion.

Detecting and removing these artefacts through traditional filtering methods is challenging due to their unpredictability and broad frequency spectrum. Therefore, the optimal approach is to prevent these artefacts through rigorous user training and the establishment of standardised procedures. By doing so, the potential for non-physiological artefacts is minimised, ensuring the reliability and integrity of EEG recordings in different experimental settings.

Physiological artefacts refer to unintended influences from various bodily processes on the recorded EEG signal, introducing complexity into its interpretation [79, 80]. These biological signals, while generally describable, exhibit high variability influenced by factors such as age, sex, and the physical-mental state. Eye-related, cardiac, and muscular signals are predominantly cited in the EEG literature, while those involving sympathetic skin responses, skin perspiration, and respiration are less common.

Eyes have a substantial influence on EEG recordings. Modelled as an electrical dipole between the retina and cornea, ocular artefacts manifest with amplitudes ranging from 50 µV to 200 µV [79, 80, 81]. The electrooculogram captures electrical activity linked to eye movements, often detected by frontal electrodes (Fp1 and Fp2). Recognisable ocular artefacts include eye blinks, producing peaks with amplitudes higher than background EEG signals (Fig. 3.1). Slow, rolling eye movements result in phase deflections related to lateral eye movements, causing alternations of opposite polarity.Notably, there is a bidirectional relationship between electrooculogram and EEG: ocular movements are both influenced by cerebral activity and contribute to measurement contamination.

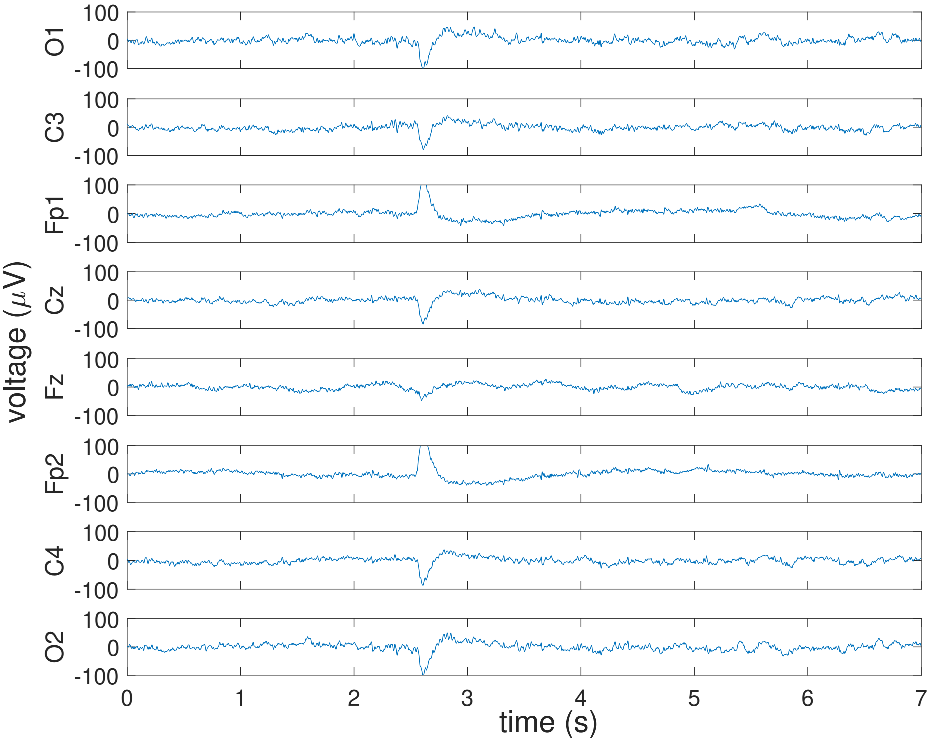


Figure 3.1: Example of eye blink in a multi-channel EEG signal.

The electrocardiogram, a record of the heart's electrical activity, exhibits lower amplitude values on the scalp compared to the chest. Recognisable patterns within the cardiac electric signal, including the P wave, QRS complex, and T wave, aid in its identification. Additionally, a cardiac pulse artefact may emerge when an electrode is positioned over a scalp artery, detecting a slow periodic wave of about 1-2 Hz (Fig. 3.2). This biomechanical artefact proves challenging to remove due to its similarity to EEG signals. Mitigation strategies involve adjusting electrode positions or employing low-pass filtering on the specific channel.

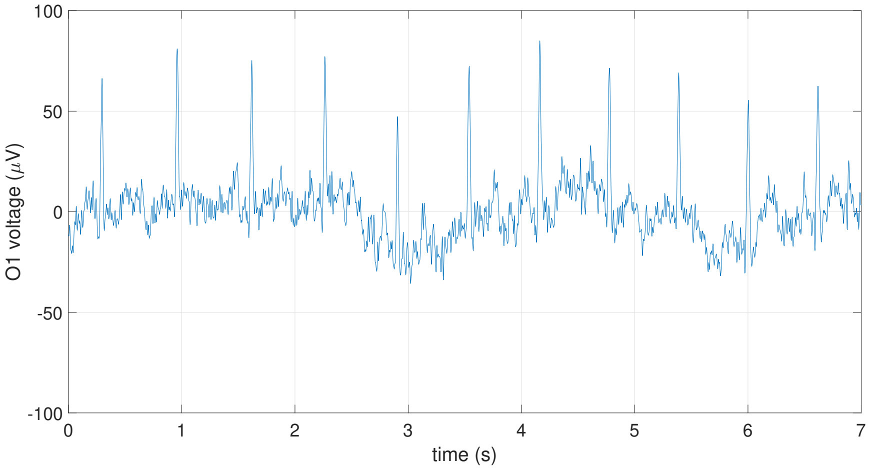


Figure 3.2: Example of cardiac artefacts on channel O1.

Muscular artefacts encompass a variety of electrical activities linked to muscle contractions and relaxations, each subject to myogenic responses controlled by the nervous system. The electromyogram captures this electrical activity, presenting irregular shapes and varying amplitudes. Spanning a frequency band of 2 Hz to 300 Hz, the electromyogram spectrum overlaps with background EEG rhythms, particularly affecting the beta rhythm from 15 Hz to 30 Hz [80]. Electromyographic artefacts exhibit a wide range of frequencies based on specific muscular activities. Actions such as chewing or swallowing showcase periodic cycles of regular bursts of potential. Facial muscle contractions and rapid eye movements also generate brief potentials detectable by the electromyogram. Notably, artefacts induced by muscles across the face, head, and neck, compounded by uncontrolled movements, significantly impact recorded signals. Experimental observations highlight the prominence of electromyographic artefacts in subjects displaying nervousness during EEG recordings, particularly among untrained users [79].

There are also other physiological sources not directly associated with ocular, cardiac, and muscle artefacts. For example, EEG electrodes placed on frontal and temporal lobes are susceptible to glossokinetic artefacts resulting from tongue movements and generic myogenic artefacts. The tongue, analogous to the eyes, acts as a bio-electric dipole generating abnormal spikes and bursts of high-frequency activity. Glossokinetic artefacts due to actions like swallowing, coughing, and chewing present oscillations with significant amplitude but varying duration. Speaking also influences EEG measurement; repetitive sounds like “lalala” result in distinctive periodic patterns on the signal [82]. Additional interference arises from respiration and perspiration, both characterised by slow, regular waves (0.25 Hz to 0.50 Hz), typically identifiable and filterable. Respiratory artefacts exhibit an almost sinusoidal trend overlapping the signal, while sweat artefacts, with less regular shapes, may lead to electrode saturation. Lastly, bone deflections, physiological abnormalities, and altered skull architecture in individual subjects can introduce unexpected rhythms, often misinterpreted as epileptic episodes. It is evident that identifying and characterising these various artefacts is crucial for accurate EEG interpretation, minimising the risk of incorrect diagnoses and ensuring the reliability of EEG information across diverse physiological contexts.

3.2 Artefact removal techniques

EEG signals pose challenges due to the presence of various artefacts, being non-stationary and non-linear stochastic mixtures. Distinguishing the actual signal of interest from artefact is crucial, giving rise to the field of artefact removal techniques. They can be grouped into four clusters [80, 83]:

- *Regression methods:* predominant until the mid-1990s, they rely on the assumption that EEG channels are linear combinations of the source signal and artefacts represented by reference channels. Despite their easy implementation and moderate computational effort, the need for reference channels and the disregard for bidirectional contamination between the signal of interest and the artefact channel limit their applicability, making them suitable for specific scenarios. Nevertheless, they are still considered the "gold standard" for method comparisons [84].

- *Filtering methods:* they play a pivotal role in signal processing [83]. Classical filters, such as low-pass and notch filters, are effective when artefact frequency bands do not overlap with the actual signal. Moreover, their main parameters, such as gain and cut-off frequency, are known. Newer techniques like adaptive, Wiener, and Bayes filtering aim to adapt parameters by optimising cost functions. Adaptive filtering assumes uncorrelated signal-of-interest and artefacts, adjusting weights during the algorithm. "Artifact subspace reconstruction" (ASR) is a recent and widely used example of adaptive technique [85], as is its Riemannian modification (rASR) [86]. Wiener filtering minimises mean square error between the actual and the estimated signal. Bayes filtering relies on probabilistic assumptions. In general, all filtering techniques have in common an a priori knowledge of the artefacts and the target system.

- *Blind Source Separation (BSS) methods:* these methods estimate a mixing matrix that connects actual sources to observed signals under certain assumptions. Techniques like Principal Component Analysis (PCA) [87], Independent Component Analysis (ICA) [88], and Canonical Correlation Analysis (CCA) [89] are commonly employed for unsupervised learning, exploiting all EEG channels simultaneously. BSS methods, while versatile, require no additional reference signal but may necessitate automatic or manual component rejection [83].

- *Source decomposition methods:*they decompose single-channel signals into multiple waveforms or units, aiming to identify and remove artefact components before signal reconstruction. Wavelet Transform [90], Empirical Mode Decomposition (EMD) [91], and non-linear mode decomposition [92] are prominent techniques in this category [83].

A significant distinction among these clusters lies in the use of channels, where BSS methods use information from all channels for transformed domain computation, while other techniques often perform independent channel-wise processing. *Hybrid methods*, combining two or more techniques, have gained attention in recent years for both multi-channel and single-channel EEG processing [80]. Additionally, the emergence of artificial intelligence and machine learning algorithms, such as neural networks and adaptive neural fuzzy inference systems, presents promising avenues for further exploration [93].

3.3 Challenges

Implementing artefact removal techniques from EEG data presents several challenges [94], particularly from a metrological perspective. The lack of standardised protocols and benchmarks in EEG analysis is a significant obstacle. Without universally agreed-upon methodologies, it is difficult to consistently compare the effectiveness of different artefact removal techniques across studies. Furthermore, the use of EEG datasets specific to researchers can introduce bias and hinder reproducibility.

Ensuring the reliability and accuracy of artefact removal algorithms is crucial from a metrological standpoint. However, selecting appropriate performance evaluation criteria remains a persistent challenge. Proposed metrics [95] often require simulated signals where both clean EEG and artefact-contaminated signals are known, making their application straightforward in such scenarios.

The most common used are: (i) mean square error that reflect the difference between the pure EEG signal and the cleaned EEG signal; (ii) the root mean square error and the normalised mean square error; (ii) signal-to-artefact ratio gain (γ) [95] that is calculated by comparing the signal-to-artefact ratio after artefact removal between the pure signal and the corrected signal, with the signal-to-artefact ratio before artefact removal between the pure signal and the contaminated signal.

However, evaluating performance in real-world applications, where only artefact-contaminated EEG data is typically accessible, becomes more challenging due to the lack of ground truth data for comparison. This limitation highlights the need for the development of novel metrics or adaptation of existing ones to accommodate real-world EEG datasets, where ground truth information may be unavailable or uncertain.

3.4 Comparative analysis of common artefact removal techniques

In the realm of wearable and portable EEG instruments, most artefact removal techniques face limitations in applicability due to their reliance on a large number of channels [93]. Moreover, there is no universally endorsed method for artefact removal, and the minimum number of channels on which these techniques demonstrate optimal performance remains unknown. Therefore, a metrological analysis was conducted to comparatively assess the main artefact removal techniques, with a specific focus on diminishing the number of EEG channels. This analysis is useful in determining the adaptability of existing techniques to scenarios involving a restricted number of EEG channels or whether novel methodologies need to be devised to address this distinct challenge. While simulated data could have been used, experimental data (i.e., real EEG) were used to enhance relevance to real-world applications. Subsequently, the following sections introduce the exploited dataset, presents the proposed analysis, and reports the obtained results.

Dataset

The EEG data were sourced from a publicly available dataset designed for evaluating artifact removal techniques [96]. It comprises recordings from 13 participants. Brain signals were acquired using a Brain Products helmet at a sampling rate of 200 Hz [97], equipped with 27 EEG electrodes and 3 EOG electrodes.

A session consists of two parts: i) two 30 s baseline acquisitions were conducted which can be regarded as a pure EEG signal as they contain few unintentional artefacts; ii) 10 repetitions from 10 s to 30 s for nine different artefact conditions were randomly presented. The artefacts included are both eye-related (i.e. blinking, closing of the eyes, rapid eye movements to the left and right and smooth eye movements to the left and right) and muscle-related (i.e. speaking, clenching the jaw and flexing the head).

In this study, blinking and closing of the eyes were processed as separate entities. Conversely, fast and smooth eye movements were consolidated into the category of eye movement artefacts due to their similarities. Similarly, all muscular artefacts were grouped together under the category of muscle artefacts. In sum, four different types of artefacts were considered. Moreover, EOG channels were not taken into account. Finally, data from subject S01 were selected for processing.

Analysis

EEGLAB [69], an open source MATLAB© toolbox for EEG processing platform, was used to remove artefacts from the data. Before applying artefact removal techniques, the EEG data were pre-processed. In particular, they were base-normalised. A bandpass filter between 1 Hz to 40 Hz was then applied.

"Next, an EEG trace of about 120 s length was obtained from the filtered data. In details, the pure data condition was entirely preserved as the first part of the trace (60 s length). For the second part, 60 s of signal with a type of artifact (eye blinking, eye closing, eye moving, or muscle artifacts) was randomly selected from the 10 repetitions and by considering a random set of channels. Therefore, a few-minute-long trace was extracted and saved as new file" [74] ©2022 IEEE.

The epoched data were processed by means of four artifact removal techniques. For each technique, the default values of the hyper-parameters were left unchanged, except for a few main parameters. In these cases, the same technique was tested with different values for these parameters chosen based on suggestions in the literature.

Each technique was tested starting with all 27 EEG channels and gradually reducing to two channels by removing one channel at each iteration. Channels were selected randomly throughout the testing process to evaluate the effectiveness of the artefact removal technique independently of specific channel choices.

The techniques exploited for this analysis were: *ASR* with two different cut-off parameters, one aggressive ($k = 15$) and one non-aggressive ($k = 25$); *rASR* with two different cut-off parameters, one aggressive ($k = 2$) and one non-aggressive ($k = 5$); *ICA* with two different rejection thresholds, one aggressive (75 %) and one non-aggressive (90 %); *PCA* with rejection of the highest variance component.

"ASR and rASR were implemented with the *clean_asr.m* function, ICA was implemented with the *runica.m* function, and PCA was implemented with the *pca.m* function.

Finally, root mean square error (RMSE) was chosen to assess artefact removal on each segment:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[EEG_{corr}(i) - EEG_{cont}(i)\right]^2}, \tag{3.1}$$

where EEG_{cont} is the contaminated signal and EEG_{corr} is the signal after one of the artifact removal techniques was applied. It must be noted that RMSE is a suggested metrics for analysis of simulated EEG data, while there is no consensus on one evaluation criterion on real EEG data [94]. However, RMSE has been used in this study to highlight the difference between the original contaminated signal and the corrected signal after the artifact removal. Ideally, an RMSE equal to zero would be desirable for the baseline condition analysis, because pure EEG epochs should be left unchanged in the artifact removal process. On the contrary, higher values of RMSE would be expected for along epochs with artifact because of the difference between contaminated EEG data and corrected EEG data.

In addition to the RMSE quantitative assessment, results were also visually inspected. For that purpose, the *vis_artifacts.m* function (included in the *clean_rawdata()* plug-in) was used to display the difference between the contaminated signal and the corrected signal.

For each subject, the analysis was repeated 10 times with random selection of channels. After these 10 runs, the mean and the standard deviation of the resulted values of RMSE were computed" [74] ©2022 IEEE.

Results

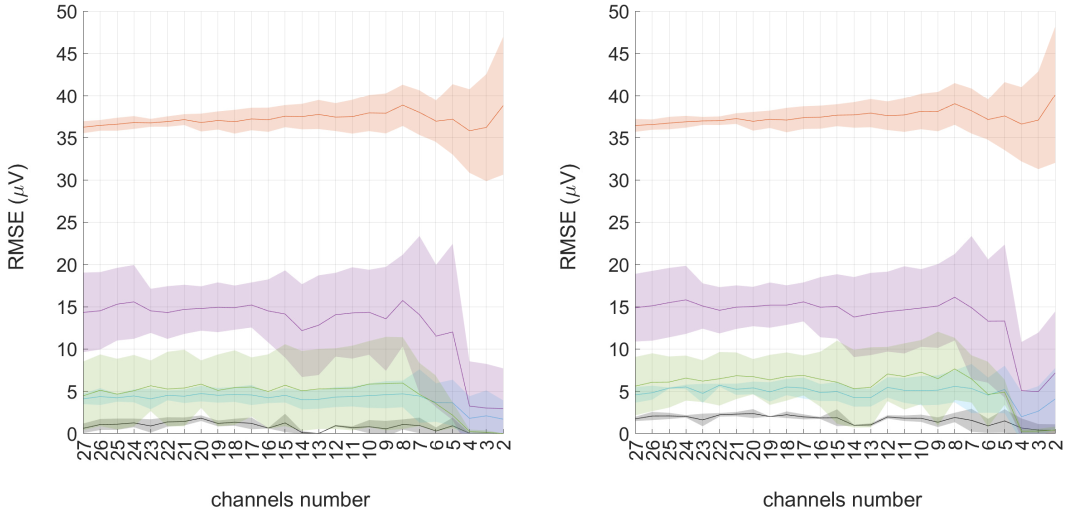
Figures 3.3, 3.4, 3.5, and 3.6 show the RMSE outcomes derived from EEG data acquired from subject s01, showcasing the impact of each artefact removal technique across varying channel configurations. The x-axis shows the number of channels used, while the y-axis shows the RMSE value in µV. Each curve on the graph illustrates the mean and standard deviation associated with specific types of EEG segments.

The black curve corresponds to the baseline EEG segment, assumed to remain constant, indicative of minimal artefact interference, thus exhibiting a trend close to zero. Conversely, the blue, green, purple, and red curves correspond to EEG segments contaminated with specific artefacts: eye closure, eye movement, eye blink, and muscle artefacts, respectively.

Artefactual segments are expected to exhibit higher RMSE values compared to artefact-free segments. This is due to the positive correlation between RMSE and artefact amplitude.

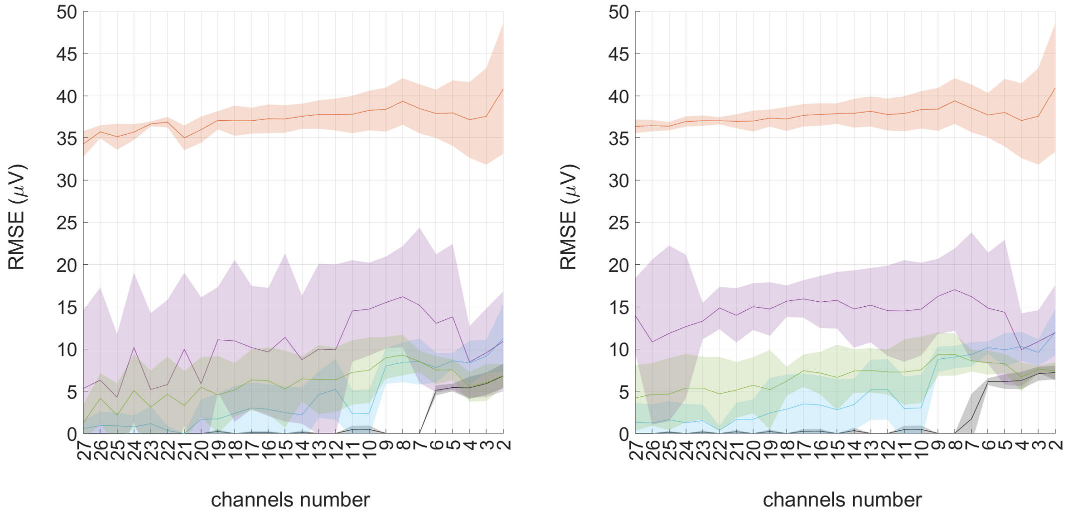
Fig. 3.3 shows the performance of ASR with a cut-off parameter of k = 25 (Fig. 3.3a) and with a cut-off parameter of k = 15 (Fig. 3.3b). The results obtained by varying the cut-off parameter are consistent. In both cases, the RMSE for the baseline condition remains below 5 µV. However, the ASR with k = 25 better preserves the baseline. ASR have been found to effectively remove all four types of artefacts even with a reduction in the number of channels used. However, a decrease in the number of channels leads to an increase in standard deviation, indicating greater performance variability. Additionally, the effectiveness of ASR decreases as the number of channels decreases, which is particularly evident when only three channels are used. ASR struggles to effectively remove eye movements (green curve) and eye closures (blue curve) in this scenario.

Fig. 3.4 shows the performance of rASR using a cut-off parameter of k = 5 (Fig. 3.4a) and k = 2 (Fig. 3.4b). The baseline remains consistent across varying channel counts, but



(a) (b)

Figure 3.3: ASR with k = 25 (a) and with k = 15 (b) applied to S01 data, respectively. Black: baseline, blue: eye closure, green: eye movement, purple: eye blink, red: muscle artefacts.

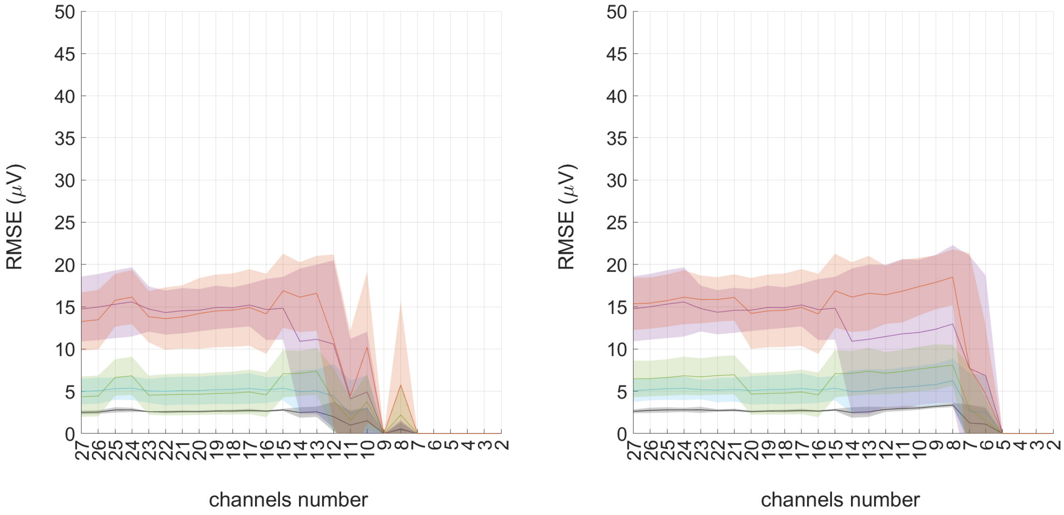


(a) rASR, k = 5. (b) rASR, k = 2.

Figure 3.4: rASR with k = 5 (a) and with k = 2 (b) applied to S01 data, respectively. Black: baseline, blue: eye closure, green: eye movement, purple: eye blink, red: muscle artefacts.

an increase in RMSE is observed from 7 channels onwards. Fluctuations in performance suggest that rASR is less stable than ASR, particularly in the removal of eye closure (blue curve) and eye blink artefacts (purple curve). Even in multi-channel scenarios, this instability persists, especially when using the less aggressive cut-off parameter. Therefore, although rASR preserves the baseline's integrity in multi-channel setups, ASR is still the preferred choice, especially in configurations with a limited number of channels.

Fig. 3.5 displays the performance of artefact removal using ICA with a less aggressive threshold (Fig. 3.5a) and a more aggressive one (Fig. 3.5b). In this case, artefact removal is less effective as the RMSE values for the different types of artefacts are lower



(a) ICA, th = 90.

(b) ICA, k = 75.

Figure 3.5: ICA with rejection threshold = 25 (a) and with rejection threshold = 15 (b) applied to S01 data, respectively. Black: baseline, blue: eye closure, green: eye movement, purple: eye blink, red: muscle artefacts.

than those obtained with ASR and rASR. According to the literature, ICA is unable to remove artefacts in a context with few EEG channels, although it is more stable than the others in a context with many EEG channels. In particular, the technique is ineffective below 9 and 5 channels when using the least and most aggressive thresholds, respectively.

Finally, the RMSE values when PCA was used as an artefact removal technique are shown in Fig. 3.6. Although all curves display a constant trend and have a very low standard deviation, the RMSE for the baseline segment is equal to or higher than 5 μV, which is higher than the other techniques. Therefore, it can be concluded that PCA also affects EEG signals without artefacts, which is confirmed by visual inspection.

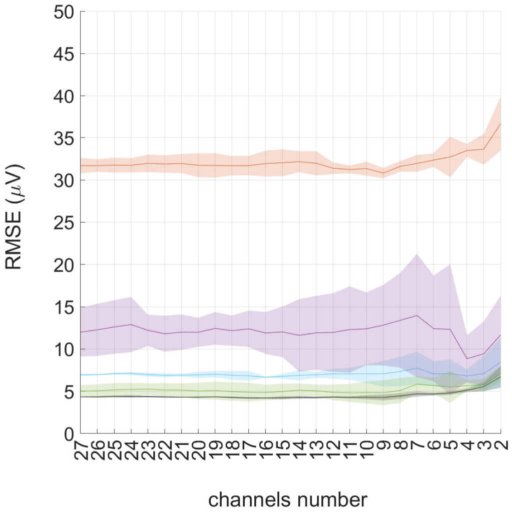


Figure 3.6: PCA applied to S01 data. Black: baseline, blue: eye closure, green: eye movement, purple: eye blink, red: muscle artefacts

In conclusion, ASR emerges as the most favourable choice for artefact removal, thanks to its stability and ability to preserve the clean EEG signal from the baseline. However, notable drawbacks are observed with rASR and PCA, as they tend to impact the clean signal. Conversely, ICA struggles to perform adequately in a setting with few channels.

As an example, Fig. 3.7 displays the EEG signals before (in blue) and after (in red) artefact removal using all four techniques in the non-aggressive version. The techniques were applied to five randomly selected channels. The figure shows only one of the five channels, Fz. The application of the techniques to the base period is compared with the removal of the eye blink segment. It is evident that only ASR and ICA do not alter the baseline, which is correct as this is an EEG segment without artefacts. As anticipated from the RMSE results, both rASR and PCA alter the baseline. However, in relation to the segment that includes eye blinks, only ASR corrected the signal accurately. Both rASR and PCA overcorrected the signal. ICA, on the other hand, faithfully follows the artefacts and is unable to remove them.

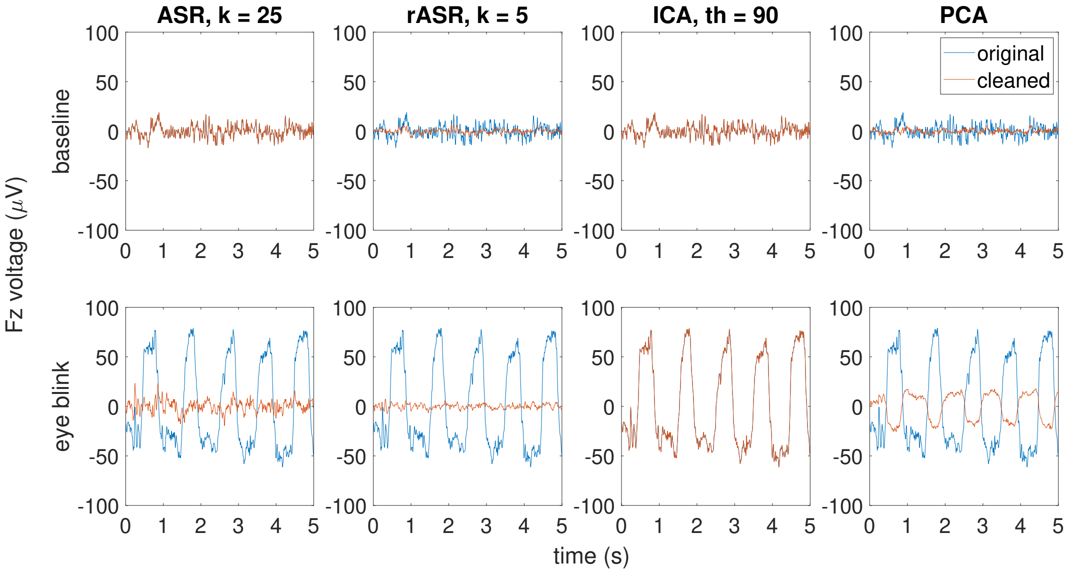


Figure 3.7: Visual inspection of the sensor Fz with five sensors involved in artifact removal by the four non-aggressive techniques. The blue curve represents the original EEG, while the red curve represents the cleaned EEG.

The EEG data from the other subjects in the dataset confirmed the results.

Several key insights can be drawn from the findings of this preliminary study. Firstly, it is evident that the RMSE related to muscle artifact segments consistently exceeds that of other artifact types for all techniques except ICA. This discrepancy can be attributed to the inherently higher noise amplitude of muscle artifacts, making their removal relatively easier for ASR, rASR, or PCA. A notable increase in standard deviation is observed as the number of channels decreases, indicating reduced performance stability. This variability in performance can also be attributed to the specific electrode locations in the few-channel setting. For example, electrodes placed on the frontal lobe tend to be more effective in removing eye-related artifacts, resulting in lower RMSE values. This observation is consistent with visual inspections conducted across different channels, which further reinforces the influence of electrode placement on artifact removal efficacy.

3.5 Novel method for artefact removal

3.5.1 Rationale

Given the challenges presented by low-density EEG setups and the limitations of current artefact removal techniques in this scenario, hybrid methods have emerged as a promising approach to improving signal quality. Hybrid methods involve integrating various rejection or removal techniques to exploit the strengths of individual methods while mitigating their respective weaknesses [80]. This approach has received attention in cases where only a small number of EEG sensors are used. In such cases, it is necessary to find innovative solutions due to the scarcity of channels.

A common hybrid approach involves a preliminary step of data decomposition to generate a higher dimensional signal. This is followed by the application of artefact removal techniques suitable for multidimensional signals. Empirical mode decomposition (EMD) [98, 99] and its variants [100] have been widely used in such hybrid methods, demonstrating improved performance when integrated with techniques such as ICA [101] or CCA [102]. In particular, recent efforts explored the combination of multivariate EMD (MEMD) with ICA [103] or CCA [104, 105] to target a specific type of artefact in low-density sensor setups.

However, while hybrid approaches show promise in improving artifact removal efficacy, they often come with increased computational costs and may focus solely on addressing a single type of artefact [102, 104, 105, 106, 103, 107, 81, 108, 109]. This limitation underscores the need for versatile techniques capable of addressing multiple artefact types simultaneously, considering the diverse nature of artefacts encountered in EEG acquisition scenarios.

In addition, the aforementioned methods often combine modal decomposition with blind source separation techniques, which rely on the strict assumption of component independence [110]. Nonetheless, these assumptions may not always be valid for modal decomposition. On the other hand, adaptive filters such as ASR [111, 86, 112] have fewer stringent assumptions and have shown to be more effective than other methods [74]. Therefore, there is a strong incentive to propose a novel hybrid artefact removal technique for low-density EEG setups. This new approach combines MEMD with ASR, allowing artefact removal from limited EEG signals, including single-sensor setups. It should be noted that ASR was chosen based on the results presented in Sec. 3.4, where it proved to be the most effective technique for removing artefacts from a limited number of EEG channels while preserving pure EEG, when compared to rASR, ICA, and PCA. However, the performance of the technique deteriorates when fewer than five EEG channels are used.

3.5.2 MEMD-ASR: methodology and implementation

In the proposed method, signals from a limited number of EEG channels are decomposed using MEMD and then ASR is applied, as is typical of hybrid approaches. MEMD decomposes the signals in the multivariate time-frequency domain, generating fundamental waveforms called intrinsic mode functions (IMFs) [113]. It differs from EMD by producing an equivalent number of scale-aligned IMFs IMFs for each signal. ASR identifies and removes artefacts in the new space defined by the IMFs resulting from the MEMD, rather than on the limited available EEG channels [111]. Finally, additional steps between MEMD and ASR are required to be effective, as a simple combination of the two is not sufficient.

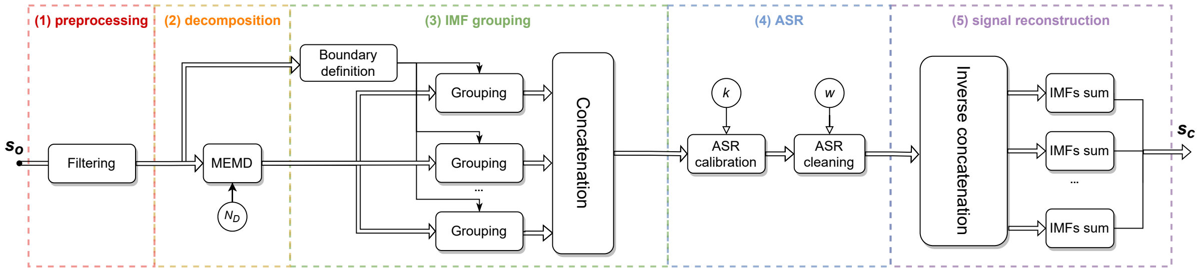


Figure 3.8: "Block diagram of the proposed hybrid method for artifact removal" [75] ©2022 IEEE. Figure from [75].

As shown in the Fig. 3.8, the technique consists of five steps:

1. "*preprocessing:* the EEG signals are filtered in the frequency bands of interest and, after that, the epochs to analyze can be selected.

2. *decomposition* the MEMD is carried out on all the available EEG signals simultaneously. The number of directions of the signal projection should be explicitly fixed. To extract meaningful IMFs, this number must be greater than the dimensionality of the original signal. The rule adopted in the current proposal is

$$N_{directions} = 2 \times (N_{sensors} + 1), \tag{3.2}$$

with the constraint of having at least six directions. This choice is consistent with the indications of the developers [114]. At the decomposition end, the IMFs of each signal are matched among the corresponding sensors.

3. *IMF grouping* the resulting IMFs have much different amplitudes, and this prevents direct application of ASR. Therefore, IMFs with low amplitudes are identified and summed to the last IMF with acceptable amplitude. To set an amplitude boundary, a calibration interval before decomposition is taken into account and the difference Δ_p between the 90$^{\text{th}}$ and 10$^{\text{th}}$ percentile is calculated for each sensor. Then, the same quantity is calculated for all the IMFs associated with that calibration interval. If the inter-quartile range of an IMF is lower than 10 % of Δ_p for at least one sensor, all the matched IMFs are summed up with the closest matched IMFs above the boundary. This process is done consistently for all sensors.

4. *ASR* it includes a calibration phase and a cleaning phase. In this hybrid method, the calibration data are the IMFs obtained from the decomposition of the calibration interval. A mixing matrix M is obtained as the square root of the covariance matrix of calibration data. PCA is carried out for this mixing matrix. Then, a rejection threshold is automatically selected for each principal component as

$$\Gamma_i = \mu_i + k\sigma_i, \tag{3.3}$$

where μ_i and σ_i are the mean and standard deviation of the root mean square values obtained with a sliding window for each component. The threshold depends on the cut-off parameter k and the time window w on which statistics are calculated. For each window, the algorithm zeroes the components exceeding the rejection limit.

5. *signal reconstruction* once the data are cleaned with the ASR, the signals in the original space are reconstructed by the inverse MEMD, i.e. the cleaned IMFs for each sensor are coherently summed up" [75] ©2022 IEEE.

To ensure proper signal cleaning, it is recommended to acquire both calibration and EEG data with artefacts under compatible and controlled measurement conditions. This includes using the same acquisition hardware, during the same session, and with the same user. However, it is possible to reuse calibration data from a different session involving the same user. Although MEMD-ASR can be extended to multidimensional signals with any number of sensors, it is important to note that the number of IMFs used in ASR increases exponentially as the number of sensors increases, resulting in longer computational time.

The method was implemented in Matlab© as a function that requires the following inputs:

1. "The original signal to be cleaned, expressed as an (S x N) matrix, where S is the number of samples and N is the number of sensors.

2. The sampling frequency of the original signal, expressed in samples per second.

3. The calibration interval corresponds to a time interval in which artefacts are assumed as absent.

4. The cut-off parameter of ASR, expressed as an integer greater than one; it determines the aggressiveness of data rejection: the smaller the k, the more the aggressiveness.

5. The length of the statistics window, expressed in seconds. This value w should be no longer than the time duration of the artifacts [111]" [75] ©2022 IEEE.

The resulting output is represented as an (S x N) matrix, containing the cleaned EEG signal. The function executes the aforementioned five steps, including the appending of an additional segment to the original signal's end. Typically, the final 0.25 s of the input signal is selected as the segment to be flipped and joined. The addition of a segment is necessary due to the default delay introduced by the ASR cleaning function, which is attributed to a lookahead operation. This option allows the ASR to reconstruct a sample by taking into account both the previous and subsequent samples [111]. The 0.25 s segment is then removed in the signal reconstruction phase downstream of the ASR.

3.5.3 Results and comparative analysis

Datasets

Both semi-synthetic and experimental datasets were exploited in order to validate the MEMD-ASR technique. The use of semi-synthetic data allows a direct comparison between pure and cleaned EEG. This allows a clear assessment of the effectiveness of artefact removal. Direct comparison is not possible with only real EEG data as there is no clean reference. Therefore, results obtained using real EEG data require careful discussion and interpretation.

Synthetic EEG signals were generated using the SEED-G simulator, a recently developed online tool for this purpose [115]. Each EEG channel generated a signal that lasted for 150 s and was sampled at a rate of 256 Hz. These parameters were carefully chosen in order to ensure compatibility with the subsequent artefact removal process by means of ASR. The first 60 s of each signal were kept free of artifacts to serve as clean calibration data. Three types of artefacts, namely muscular, ocular and blinking, were introduced during the remaining 90 s. The DenoiseNet dataset [116] was used to derive ocular and muscle artefacts that are available in 2 s segments at the same sampling frequency as the synthetic pure EEG signal. For the purposes of this work, 15 segments with muscle artefacts and 15 with ocular artefacts were randomly extracted, resulting in a total of 30 s of artefacts per each artefact type. Blink artefact was also considered. These artefacts were simulated using half-sinusoidal signals lasting 0.2 s, spaced 1 s apart to mimic typical eye-blinking patterns. These occurred during the last 30 s of the signal. To ensure a realistic simulation, the amplitude of the artefacts was corrected with respect to the pure synthetic signal, maintaining a signal-to-noise ratio of −20 dB to 5 dB [117].

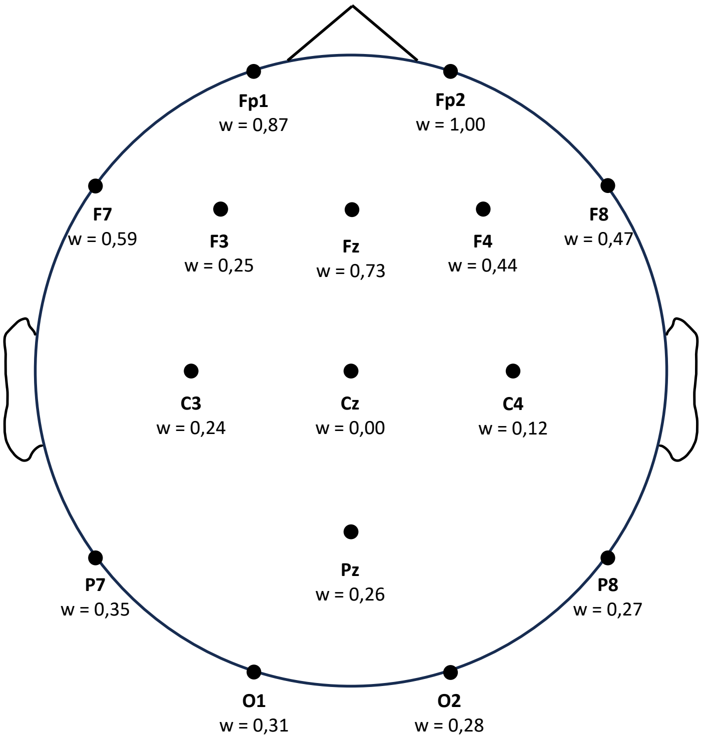


Figure 3.9: Electrode placement and correlation coefficients (w) associated with each electrode to weight ocular artefacts and eye blinks.

Special attention was given to the propagation of ocular and eye blink artefacts, which typically originate from the prefrontal area of the scalp. The inclusion of these artefacts was weighted for each of the 15 channels selected from various regions of the scalp (see Fig. 3.9). The weighting process was based on correlation coefficients obtained by comparing actual electrooculographic data with corresponding EEG data [115]. For experimental data, the dataset described in Sec. 3.4 was used.

Experimental setup

The proposed MEMD-ASR underwent testing using data acquired from four EEG electrodes down to a single EEG electrode, demonstrating effective operation with up to four EEG electrodes. It is worth noting that the MEMD reverts to EMD when operating with a single EEG electrode. To ensure an unbiased assessment of performance regardless of EEG electrode location, a random selection of EEG electrodes was made from the available pool for each EEG electrode configuration. To identify the optimal pairing, various values for the main parameters of ASR, namely k and w, were explored. Specifically, the optimal value of k was chosen in a range from 5 to 30, with a step of 1, while the optimal value of w was chosen in a range from 0.2 s to 3.0 s with a step of 0.1 s. The choice of the best values was guided by the Relative Root Mean Square Error (RRMSE) [95], which quantifies the differences between the pure EEG signal and the signal cleaned by MEMD-ASR as follows:

$$RRMSE = \sqrt{\frac{\sum_{j=1}^{N}\left[EEG^{*}(j) - EEG(j)\right]^{2}}{\sum_{j=1}^{N} EEG(j)^{2}}}, \tag{3.4}$$

where $EEG(j)$ corresponds to the j-th pure EEG sample and $EEG^{*}(j)$ are samples of the pure signal and cleaned one, respectively. RRMSE values approaching zero indicate successful artifact removal. Notably, Eq. 3.4 can only be applied when semi-synthetic EEG data are available, as it represents the sole scenario wherein the pure and uncontaminated EEG signal can be obtained. However, in this study, RRMSE will also be computed using experimental EEG data, albeit restricted to baseline segments and artifact-free EEG segments. In these instances alone, RRMSE will carry the same meaning.

Using semi-synthetic data, the RRMSE was assessed on various segments of the EEG signal, including the baseline segment used for the ASR calibration, residual pure EEG segments, and segments contaminated by different types of artefacts (i.e. muscle artefacts, eye movements, and eye blinks). "The point defined by the five RRMSE values (as five were the EEG segments) was considered in a Euclidean 5-D space and its norm was calculated. The dispersion around these values was also taken into account and propagated to achieve the uncertainty of the norm. The optimal w was chosen to minimize the Euclidean norm and possibly its associated uncertainty. Then, k was identified by looking for the minimum 90th percentile associated with the five values. As an essential check, once w and k were fixed, a visual inspection of the signals in the time domain was carried on. In particular, the pure signal and the signal cleaned with MEMD-ASR were compared for each segment of the EEG signal" [75] ©2022 IEEE. The hybrid method was then validated with experimental data exploiting the best w and k parameters determined from semi-synthetic data. To evaluate the performance of different sensor configurations, sensors were randomly selected as for semi-synthetic data. In this case, the effectiveness of MEMD-ASR was assessed through visual inspection and RRMSE calculation, with a focus on baseline and pure EEG segments. RRMSE values for segments contaminated by artefacts were not reported to avoid misinterpretation, as there was no pure signal in the experimental data.

For comparison purposes, the results obtained with the proposed MEMD-ASR were compared with those obtained with ASR and MEMD alone. Furthermore, ASR was applied up to two EEG electrodes, as it is a covariance-based method [53, 118]. Therefore, in the case of a single EEG electrode, EMD-ASR was only be compared with EMD. Finally, IMFs containing artefacts were identified and removed by means of entropy, as proposed in [119].

Results on semi-synthetic data

Fig. 3.10 reports an example of Euclidean norm in 5-D space associated with the medians of the RRMSEs calculated on the five EEG segments, when using two EEG electrodes. The colours reflects the dispersion around the median RRMSEs values. Similar surfaces were obtained in the other cases.

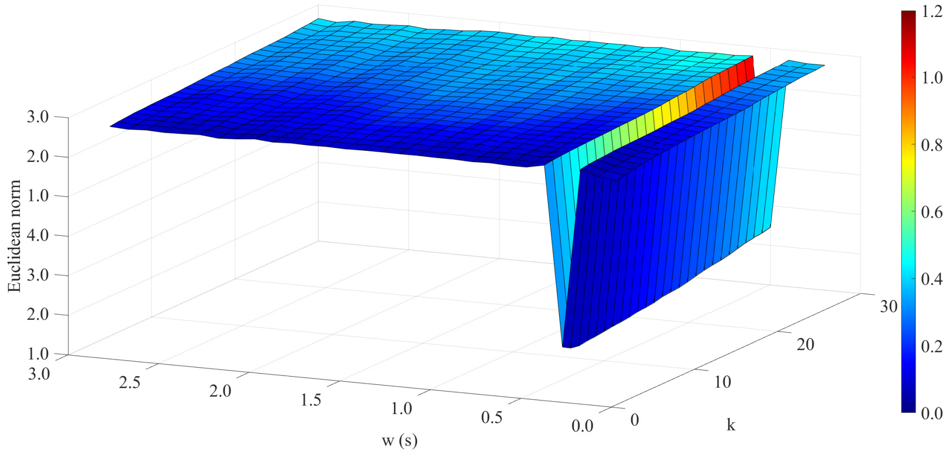


Figure 3.10: Euclidean 5-D norm associated with median RRMSEs, when using two EEG electrodes. The colours reflect the dispersion of the norm.

The optimal value of w, which is equal to 0.5 at the negative peak, can be identified from the surface. This value remains constant regardless of the k-value and the number of EEG electrodes used. This finding is consistent with the default length of the time window w on which statistics are calculated.

Table 3.1: "Median RRMSE values for semi-synthetic data" [75] ©2022 IEEE. Table from [75].

#sensors	method	w (s)	k	baseline	pure	muscle	ocular	blinks
4	MEMD	-	-	0.00	0.00	8.61	0.99	0.56
4	ASR	0.5	7	0.00	0.00	0.27	0.34	0.40
4	MEMD-ASR	0.5	9	0.00	0.00	0.60	0.30	0.46
3	MEMD	-	-	0.18	0.20	8.87	0.95	0.59
3	ASR	0.5	8	0.00	0.00	0.38	0.37	0.47
3	MEMD-ASR	0.5	8	0.01	0.00	0.70	0.34	0.45
2	MEMD	-	-	0.25	0.27	9.08	0.90	0.54
2	ASR	0.5	12	0.00	0.00	0.69	0.58	0.55
2	MEMD-ASR	0.5	9	0.00	0.00	0.70	**0.35**	0.45
1	EMD	-	-	0.27	0.31	9.24	0.65	0.62
1	EMD-ASR	0.5	12	0.00	0.00	1.25	0.50	0.59

"**bold**: median RRMSE for MEMD-ASR significantly lower than respective ASR one."
"w: length of the statistics window in seconds; k: cut-off parameter of ASR."

Moreover, the surface indicates that the optimal value of k should be selected from the range of 5 to 10. In this range, k minimises both the norm and its dispersion. The optimal k-value for varying the number of EEG electrodes was calculated as reported in Sec. 3.5.3. In general, when using small values of k, the technique effectively removes artefacts but also alters the pure EEG signal. Conversely, when using large values of k, the artefactual segments are only partially cleaned, while the pure EEG signal remains unchanged.

Tab. 3.1 shows the medians of the RRMSE values using the optimum w and k for different numbers of EEG electrodes. The values are shown for MEMD-ASR, for ASR alone and for MEMD alone. The performance of the MEMD-ASR was similar to that of the ASR, while the MEMD alone failed to clean the signal from artefacts and distorted the clean signal. This is due to its filter bank-like behaviour [113]. The obtained results are consistent with those reported in the literature, where the RRMSEs range from 0.2 to 0.5 when simulating EEG data with an SNR in range −20 dB to 5 dB [104, 105, 106]. It is worth noting that all studies refer to a minimum number of EEG channels of three [103, 114].

To assess significant differences in efficacy between MEMD-ASR, ASR and MEMD, the Mann-Whitney U-test was used [120]. The significance level was set at 5 %. The test proved as well as visual inspection that: (i) MEMD-ASR and ASR are compatible, except for ocular artefacts with two sensors (shown in bold in Tab. 3.1); (ii) MEMD performs worse than both MEMD-ASR and ASR.

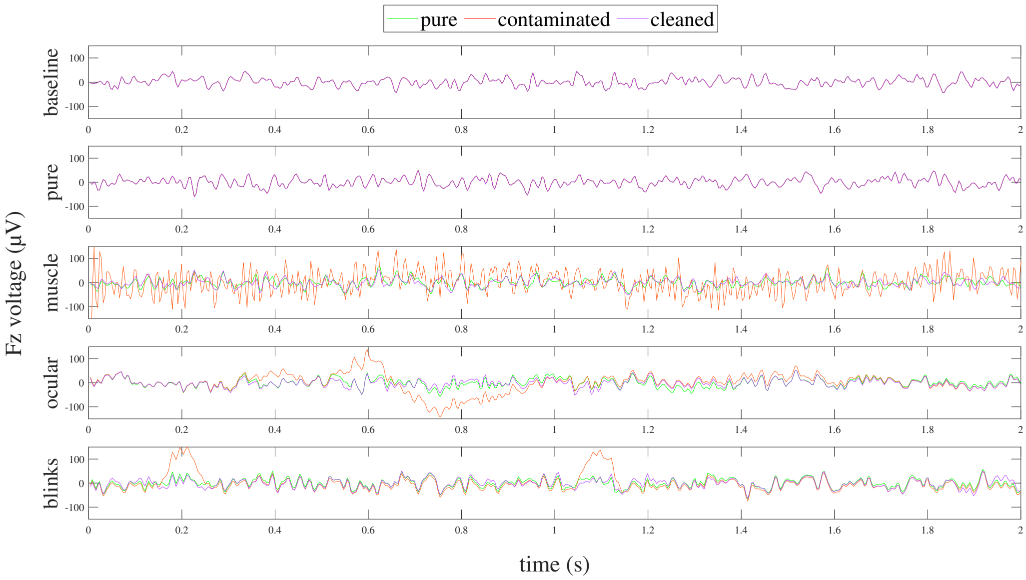


Figure 3.11: Semi-synthetic data: visual inspection of Fz electrode when MEMD-ASR was applyed on two electrods.

Fig. 3.11 presents a representative example of visual inspection in the two-electrode case. For each of the five segments, electrode Fz was reported. Each plot displays three curves: the pure EEG signal (green curve), the contaminated signal (orange curve), and the signal cleaned with MEMD-ASR (purple curve). The pure and cleaned signals are mostly overlapped, demonstrating the effectiveness of the proposed technique.

Results on experimental data

Due to the lack of the pure EEG signal, the optimum values of w and k chosen using the semi-synthetic data were also used with the experimental data.

Table 3.2: "Median RRMSE for experimental data" [75] ©2022 IEEE. Table from [75].

#sensors	method	w (s)	k	baseline	pure
4	MEMD	-	-	0.94	0.94
4	ASR	0.5	7	0.90	0.90
4	MEMD-ASR	0.5	9	**0.32**	**0.28**
3	MEMD	-	-	1.01	1.01
3	ASR	0.5	8	1.14	1.14
3	MEMD-ASR	0.5	8	**0.34**	**0.40**
2	MEMD	-	-	0.80	0.75
2	ASR	0.5	12	0.94	0.98
2	MEMD-ASR	0.5	9	**0.19**	**0.28**
1	EMD	-	-	0.72	0.67
1	EMD-ASR	0.5	12	0.18	0.15

"**bold**: median RRMSE for MEMD-ASR significantly lower than respective ASR."
"w: length of the statistics window in seconds; k: cut-off parameter of ASR."

Tab. 3.2 shows the median RRMSE values for different numbers of EEG electrodes. Again, the values are shown for MEMD-ASR, for ASR alone and for MEMD alone. In contrast to the previous case, results were only reported for baseline and pure EEG segments. In general, MEMD-ASR affects artefact-free EEG segments less than ASR and MEMD alone. The Mann–Whitney U-test confirmed a significant difference for all cases from four to two EEG electrodes.

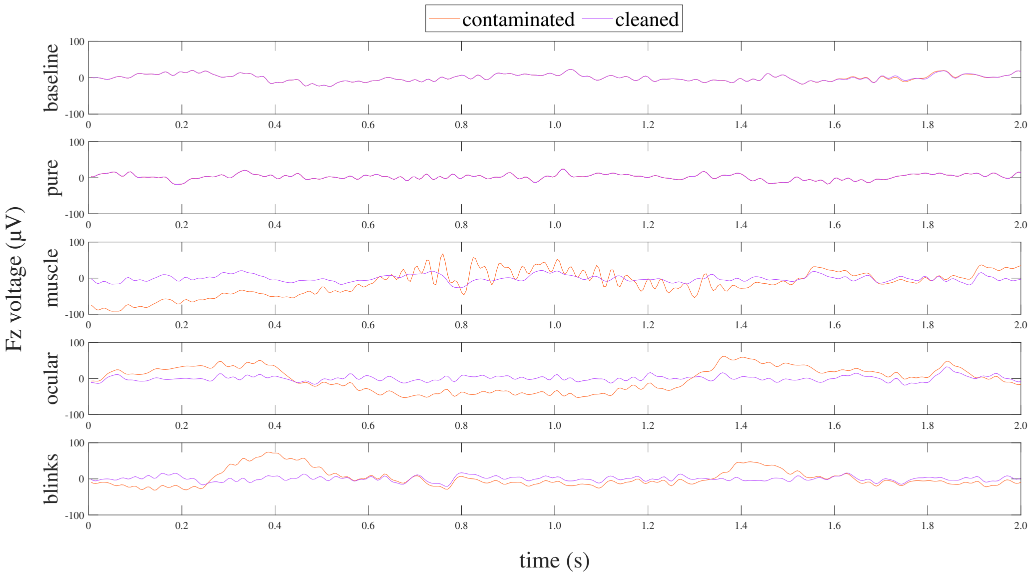


Figure 3.12: Experimental data: visual inspection of Fz electrode when MEMD-ASR was applyed on two electrods.

Fig. 3.12 presents a representative example of visual inspection for two electrodes using experimental data. The signal measured by the Fz electrode is plotted for each of the five segments. Each graph shows two curves: the experimental data (orange curve) and the signal cleaned with MEMD-ASR (purple curve). The curves mostly overlap for the baseline and pure segments, while muscle, ocular, and blink artefacts were correctly removed.

Discussion

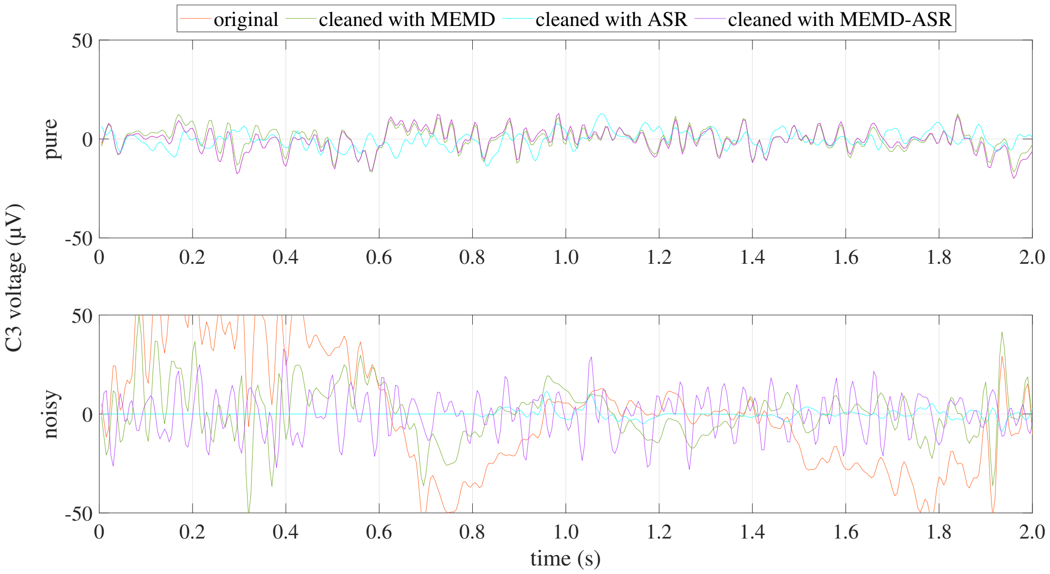


Figure 3.13: Experimental data: visual inspection of C3 electrode when MEMD-ASR, MEMD, and ASR were applied on two electrods.

The MEMD-ASR technique was tested on both semi-synthetic and experimental data. Simulated data are commonly used by researchers to perform preliminary analysis on a method of artefact removal of interest [94] but they are not a true full representation of actual EEG data. However, as the pure EEG signal is available in simulated data, objective metrics can be applied to evaluate the effectiveness of a proposed technique. However, currently there is no metric capable of accurately measuring the effectiveness of this technique on experimental data. The RRMSE was used to measure the performance of the hybrid technique proposed, as well as ASR and MEMD, on both semi-synthetic and experimental data. The results showed that ASR performs similarly or even better than MEMD-ASR on semi-synthetic data down to two sensors, while MEMD-ASR consistently outperforms on experimental data. As already known in the literature [113], MEMD alone is not effective in cleaning the EEG signal.

However, upon visually inspecting the removal of artefacts from experimental signals using two EEG electrodes (Fig. 3.13), neither MEMD nor MEMD-ASR affected the signal during the pure segment. In contrast, ASR perturbed it. With respect to the noisy segment, only MEMD-ASR was capable of effectively removing ocular artefacts. Conversely, MEMD failed to mitigate artefacts, and ASR overly corrected them. In contrast to the results on semi-synthetic data, ASR did not remove artefacts and cancelled out the signal in experimental data. Therefore, despite efforts to identify a reliable metric, it is still necessary to rely on visual inspection as the RRMSE values did not indicate the failure of ASR on experimental data.

Finally, it is important to note that the proposed technique has a drawback in terms of its execution time. Specifically, the execution time for ASR alone ranges from approximately 0.2 s to 0.4 s for artefact removal. The time increases with more EEG electrodes.

In contrast, MEMD-ASR execution time ranges from approximately 32 s (two EEG electrodes) to about 72 s (four EEG electrodes). Additionally, EMD-ASR for a single electrode requires approximately 0.7 s. Therefore, MEMD creates a bottleneck in processing time. Instead, EMD-ASR offers the potential to implement a soft EEG electrode with integrated artefact removal capabilities.

Chapter 4

Signal processing for motor imagery brain-computer interfaces

In the context of brain-computer interfaces, once the EEG signal has been acquired and corrected for artefacts, signal processing is required for the system to properly work. In this chapter, a detailed metrological analysis of recent EEG signal processing techniques is carried out. It concerns the performance of the various algorithms used for feature extraction, classification, and translation of motor imagery signals into control commands. The reliability of these algorithms is crucial, as they constitute a fundamental component of BCI functionality.

Some of the work described in this chapter has been previously published in [16].

4.1 Metrological analysis of literature

This section presents how the review of the existing literature was carried out. The aim is to identify and compare the most effective processing strategies for the classification of EEG signals. To this end, a standardised procedure was used to select the studies. An in-depth performance analysis was conducted using the international metrology vocabulary. Statistical tools were used for an objective assessment of classification performance.

4.1.1 Review of existing processing strategies

This systematic review was carried out in accordance with the "Preferred Reporting Items for Systematic Reviews and Meta-Analyses" (PRISMA) guidelines [121], as showed in Fig 4.1.

The literature search focused on studies addressing the classification of MI signals in neural interfaces, explicitly excluding those focusing on channel reduction or training enhancement strategies. For comparison purposes, only studies using at least one public dataset were screened. This screening criterion ensures that the proposed processing methods can be replicated in the same experimental settings to verify the consistency of the results. By testing the proposed processing methods on different datasets (both public and non-public), the reproducibility of the results can be also assessed. The search was conducted using Scopus and PubMed as search engines. For both, were used the keywords:

- BCI, motor imagery, public dataset;

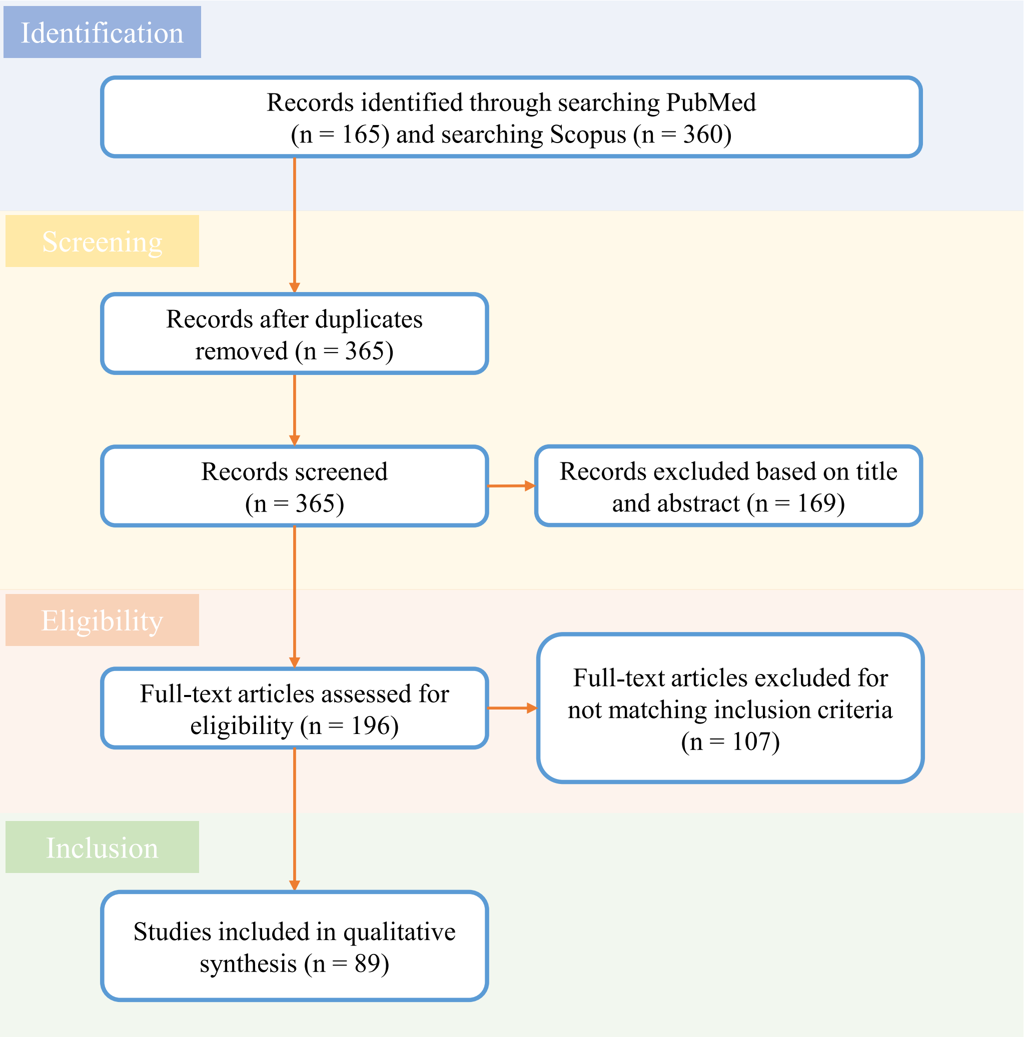


Figure 4.1: The PRISMA flow chart illustrates the number of studies identified, screened, elected, and included in this literature review.

- BCI, competition, dataset;
- motor imagery, dataset.

The term BCI was used both in extended form and as an acronym. All English-language studies published since 2017 and for the following 4 years were identified.

Duplicate studies were excluded. Then, studies were discarded by using the information in the title and abstract of each paper if:

- they used brain signals acquired by techniques other than EEG (or EEG in combination with other techniques);
- they did not use the entire dataset, but selected subjects within it, or used data from a single subject;
- they had main objectives other than EEG signal classification (e.g. reducing the number of EEG channels, or reducing training time).

After this initial screening, the full texts of the remaining articles were evaluated for eligibility, leading to the exclusion of additional papers that did not meet the criteria. In the end, 89 studies in total were included in this review.

The following information was extracted from the included studies: (i) dataset, (ii) number of MI tasks taken into account, (iii) processing strategy, (iv) assessment method, and (v) performance metric. For each study, detailed information about the dataset was extracted. This included whether the dataset was publicly available, the name of the dataset, the number of subjects involved, the number of EEG channels, and the specific types of motor imagery tasks performed. In cases where the dataset was not public, additional details were gathered, such as the EEG instrument used for data acquisition, the type of electrodes, the number of channels, reference and ground channels, the sampling rate, any filters applied, the number of sessions and trials conducted, and the specific MI tasks undertaken.

The studies were analysed based on the results reported in their respective publications. A rigorous metrological analysis was carried out by notably considering a set of statistical tools for objective quantification. Key aspects of concern included the uncertainty, repeatability, and reproducibility of the results. In accordance with the "international vocabulary of metrology" [122]: *uncertainty* is a "non-negative parameter characterising the dispersion of the quantity values being attributed to a measurand, based on the information used", *repeatability* is the "closeness of the agreement between the results of successive measurements of the same measurand carried out under the same conditions of measurement" (repeatability conditions), and *reproducibility* is the "closeness of the agreement between the results of measurements of the same measurand carried out under changed conditions of measurement" [122].

"By exploiting a type-A statistical assessment [123], the uncertainty of a mean performance value was calculated as the associated standard deviation divided by the square root of the number of values contributing to the mean. In the present analysis, the standard deviation was either retrieved from each study, or it was calculated from the reported performance values. In calculating the repeatability, the range of performance values was considered once the dataset was fixed. This range corresponded to the difference between the maximum value and the minimum one. Finally, the same criterion was used for calculating the range associated with the reproducibility, though in this case the performance values were associated with different datasets" [16].

To ensure a comprehensive comparison, only studies using at least one public dataset were considered. Focus was placed on metrics for reporting MI-BCI performance before identifying the best proposals. Statistically-based analyses considered the distribution of reported performance values, identifying not only median performances but also the top-performing approaches with results above the upper quartile. This focus on promising approaches highlighted successful processing strategies and emerging trends in the field. The results of these analyses are reported and discussed in the following sections, providing insights into the current state-of-the-art in MI-BCI signal processing and guiding future research towards more effective and reliable BCI systems.

4.2 Identification of successful approaches

This section presents the outcomes of the analysis and it focuses on the used data, the employed processing methods, performance evaluation techniques, and the achieved results. Emphasis was placed on identifying common trends in these studies, tracing their chronological development, and offering insights on how to create effective processing strategies for motor imagery classification. Particular attention is given to metrological aspects, with the aim of rigorously quantifying current performance levels and assessing the potential for replicating the most promising findings.

4.2.1 Datasets

A total of twelve public datasets were used in the studies under review, including those provided by various BCI competitions. A concise overview of the public datasets is given below, detailing the number of participants, the number of EEG channels, the types of MI tasks and the percentage of use derived from the included studies:

- *"BCI Competition IV - Dataset 2a"* [124]:
    - *Participants*: 9
    - *EEG channels*: 22 EEG electrodes placed over the sensorimotor area, along with 3 electrooculographic (EOG) electrodes
    - *MI tasks*: Imagination of left hand, right hand, feet, and tongue movements
    - *Usage*: 58 %
- *"BCI Competition IV - Dataset 2b"* [125]:
    - *Participants*: 9
    - *EEG channels*: 3 bipolar EEG electrodes (C3, Cz, and C4) and 3 monopolar EOG electrodes
    - *MI tasks*: Imagination of left hand and right hand movements
    - *Usage*: 40 %
- *"BCI competition III - Dataset 4a"* [126]:
    - *Participants*: 5
    - *EEG channels*: 118
    - *MI tasks*: Imagination of right hand and right foot movements
    - *Usage*: 30 %
- *"BCI competition III - Dataset 3a"* [126]:
    - *Participants*: 3
    - *EEG Channels*: 64 electrodes
    - *MI Tasks*: Imagination of left hand, right hand, foot, and tongue movements
    - *Usage*: 19 %
- *"BCI Competition IV - Dataset 1"* [127]:
    - *Participants*: 7 (including 3 artificially generated participants)
    - *EEG channels*: 59 electrodes placed over the sensorimotor area
    - *MI tasks*: Two classes selected from left hand, right hand, and foot movements.
    - *Usage*: Used in 15 % of the cases.

- *“Cho2017 dataset”* [128]:
  - *Participants*: 52
  - *EEG channels*: 64 electrodes
  - *MI tasks*: Imagination of left hand and right hand movements
  - *Usage*: Used in 3 % of the cases.
- *“High-Gamma Dataset”* [129]:
  - *Participants*: 14
  - *EEG channels*: 128 channels
  - *MI tasks*: Imagination of left hand, right hand, both feet movements, and a rest condition
  - *Usage*: Used in 3 % of the cases
- *“PhysionetMI Dataset”* [130]:
  - *Participants*: 109
  - *EEG channels*: 64 electrodes
  - *MI tasks*: Imagination of opening and closing of left/right fists and both fists/feet
  - *Usage*: Used in 2 % of the cases
- *“MAMEM Phase I Dataset”* [131]:
  - *Participants*: 34 (18 able-bodied and 16 motor-impaired)
  - *EEG channels*: 61 electrodes
  - *MI tasks*: Imagination of right hand and left hand movements
  - *Usage*: Used in 2 % of the cases
- *“Upper limb movements Dataset (ULM)”* [68]:
  - *Participants*: 15
  - *EEG channels*: 61 electrodes
  - *MI tasks*: Kinesthetic imagination of six movements, including elbow flexion/extension, forearm supination/pronation, and hand open/close
  - *Usage*: Used in 1 % of the cases
- *“SMR-BCI dataset”* [132]:
  - *Participants*: 14
  - *EEG channels*: 15 electrodes
  - *MI tasks*: Imagination of right hand and feet movements
  - *Usage*: Used in 1 % of the cases
- *“SJTU dataset”* [133]:

  - *Participants*: 5
  - *EEG channels*: 62 electrodes
  - *MI tasks*: Imagination of left and right hand movements
  - *Usage*: Used in 1 % of the cases

As highlighted by the percentage usage frequency, the public datasets were employed to varying degrees. For example, the "BCI competition IV, 2a" dataset [124] was utilised in 58 % of the studies examined, which serves to illustrate its popularity and reliability within the field. This dataset comprises data from nine participants whose EEG signal was acquired via 22 channels. The participants were instructed to imagine movements of the left hand, right hand, both feet and tongue. In contrast, only 1 % of the studies used datasets like the one from Shanghai Jiao Tong University "SJTU" [133], despite being publicly accessible. This indicates a lack of availability or lower perceived relevance in the field. As a matter of fact, although "SJTU" was declared a public dataset by the authors, it seems to be unavailable on any platform.

In many instances, researchers used multiple datasets to validate their proposed methods. However, it is notable that only a small fraction of the studies (namely 9 of 89) employed proprietary, non-public datasets to further test their proposed methods. This shows a significant dependence on publicly available datasets within the field. While this is beneficial for standardisation, it may limit the exploration of more varied data sources.

In order to ensure the repeatability of results, Tab. 4.1 provides detailed descriptions of the non-public datasets. This encompasses information on subjects, instruments, electrode types, channels, reference and ground channels, sample rate, filters applied, and the number of sessions and trials. Furthermore, the tasks performed in these datasets are specified, thereby providing a comprehensive understanding of the conditions under which the data were collected.

It is important to note that some gaps exist in the information provided for the non-public datasets, as certain details were not available in the original studies. For instance, many of the reviewed works did not specify the type of electrodes used to acquire the EEG data. Such omissions underscore the need for greater standardisation in reporting these details, as they are essential for replicating results and for ensuring the comparability of different studies.

Across both public and non-public datasets, the majority relied on synchronous acquisition paradigms. The exceptions were "BCI Competition IV - Dataset 1" and online experiments referenced in studies [140, 138], which employed self-paced (i.e. asynchronous) paradigms. This distinction is important as it may influence the processing strategies and the resulting classification performance.

4.2.2 Common processing approaches

Acquired EEG signals often contain various artefacts from both external (e.g. environmental noise) and internal (e.g. physiological processes) sources. As discussed in Chapter 3, these artefacts can impact the quality of the signals. Consequently, a pre-processing phase is often required. This step is of particular importance in online applications [93], where clean signals are required for accurate MI classification

The most commonly used pre-processing methods in MI-BCI systems are [48]:

- *Temporal filtering*: Butterworth and Chebyshev filters are typically used to filter the EEG signal in the frequency range between 8 Hz and 30 Hz. These filters are

	Subjects	Device	Electrode type	Channels	Reference channel	Ground channel	Sample rate (Hz)	Filters	Sessions and trials	Tasks
[134]	12			64			250	8-30 Hz BP	234 training trials, 234 testing trials	LH, RH
[135]	12	Neuroscan		60	2 (n.s.)		1000			LH, RH
[136]	11	eegoTMrt Ant Neuro		64	CPz	AFz	512	0.5-40 Hz BP	160 trials	dominant hand MI (kinesthetic grasping)
[137]	8	Open BCI	dry	8	Fz	A2			5 sessions, 120 trials per session	LH, RH, LF, RF
[138]	8	UE-16B EEG amplifier		16	A1, A2	forehead	1000	100 Hz LP, 50 Hz notch	1 training session (100 trials), 1 online session	LF, RH, F, T
[139]	5	g.tec		2	Fz		256	50 Hz notch	2 sessions	LH, RH
[140]	5	Emotiv Epoc+	semi dry	6	P3 and P4		128		180 offline trials, 420 online trials	LH, RH
[141]	5	SynAmps2 system (Neuroscan)		21	vertex	forehead	1000	0.5–200 Hz BP	3 sessions, 40 trials per session	RH index finger (extensions/ flexions), idle state
[141]	4	SynAmps2 system (Neuroscan)		27	mastoid behind the left ear	forehead	1000	0.5–200 Hz BP	3 sessions, 40 trials per session	LH, RH
[142]	1	Emotiv Epoc					128		160 trials	LH, RH

Table 4.1: "Details about the non-public datasets employed by the studies included in the comparative analysis. Empty spaces are associated with unavailable information. Channel position refers to the 10-20 or 10-10 standards. BP: band-pass, LP: low-pass. LH: left hand, RH: right hand, F: both feet, LF: left foot, RF: right foot, T: tongue, (n.s.): not specified" [16].

effective in removing unwanted noise and preserving relevant information in the μ and β frequency bands, which are crucial for MI tasks.

- *Spatial filtering*: common reference average and Laplacian filters [143] are often used to reduce common noise and improve the spatial resolution of EEG signals by isolating activity from specific brain regions associated with motor functions.

The next phase is feature extraction, in which various methodologies are employed to capture the most relevant information from the pre-processed EEG signals. Feature extraction can be broadly classified into three main domains:

- *Time-domain methods*: They analyse the sensorimotor rhythms over time, using statistical measures such as mean, variance, skewness, and kurtosis [28]. Other approaches include power analysis, auto-regressive (AR) modelling [144], and graph-based techniques [145], which typically operate on a channel-by-channel basis.

- *Frequency-domain methods*: They focus on identifying dominant frequencies within the EEG signals, using methods like “fast Fourier transform” (FFT) [146], “power spectral density” (PSD) [147], and band power estimation [148]. These approaches help in capturing the frequency characteristics relevant to MI tasks.

- *Time-frequency methods*: As temporal or frequency features may not fully capture the necessary information, time-frequency methods, such as the “short-term Fourier transform” (STFT) [149] and wavelet transforms [150], are frequently employed. These techniques offer a more detailed analysis by examining both temporal and spectral characteristics simultaneously.

- *Spatial-domain methods*: Among spatial-domain techniques, “common spatial pattern” (CSP) and its variants are the most prominent [7]. CSP is often combined with filter banks to extract band-specific spatial features [151]. “Independent component analysis” (ICA) and Laplacian filtering are also extensively used to enhance the spatial resolution of the signals [7].

Following the extraction of features, a selection process is required to remove any redundant or irrelevant information, thus ensuring that only the most informative features are employed in the subsequent classification stage. The most commonly employed methods for feature selection include correlation criteria and mutual information [48].

Finally, classification strategies are employed to categorise the EEG signals into different MI tasks. Traditional classifiers like decision trees [152], “support vector machines” (SVMs) [153], and “linear discriminant analysis” (LDA) [154] have been widely used. However, recent advancements have seen a growing focus on “deep neural networks” (DNNs) [155] and “spiking neural networks” (SNNs) [156], which offer improved classification performance due to their ability to model complex patterns in the data. It is notable that these recent approaches have the potential to integrate feature extraction, feature selection and classification into a unified pipeline.

Despite their advantages, deep learning models are prone to overfitting, especially given the limited amount of available EEG data [157]. To address this challenge, techniques such as data augmentation and transfer learning are often employed. Data augmentation involves artificially expanding the training dataset by generating new samples from existing data, which enhances the model’s generalisation capability [158]. On the other hand, transfer learning leverages pre-trained models [157], reducing the amount of training data required and speeding up the training process.

Consistent with the existing literature [159, 160, 161], this analysis shows that EEG signal processing methods in MI-BCI systems can be classified into:

- “non-brain-inspired” (non-BI) approaches: they rely on traditional machine learning algorithms that learn patterns from data without explicit reference to how the brain processes information. These approaches typically involve feature extraction using pre-defined methods followed by standard machine learning classifiers, such as LDA or SVMs.

- “brain-inspired” (BI) approaches: they draw inspiration from the structure and function of the human brain, particularly neurons. This category includes “artificial neural networks” (ANNs), DNNs, and SNNs. These models simulate the way neurons process information, using nodes that receive inputs, apply weights and biases, and produce outputs based on learned activation functions. ANNs and DNNs mimic synaptic scaling, where each node represents a neuron that integrates input signals and applies a nonlinear transformation. DNNs are a subset of ANNs characterised by multiple hidden layers that enable the modelling of complex patterns [160]. On the other hand, SNNs are designed to emulate the spiking behaviour of biological neurons, transmitting information through discrete pulses or "spikes", where both the amplitude and timing of these spikes carry critical information [156].

A trend analysis of EEG signal processing approaches for MI-BCI systems reveals a change in methodology over the period from 2017 to 2021. During this period, there has been a move away from traditional non-BI methods in favour of BI approaches. Specifically, the analysis shows that non-BI methods were employed in about 53 % of the studies, while BI approaches were used in about 47 % of the cases. However, a closer examination of the chronology indicates that non-BI methods have experienced a gradual decline in use over the time period considered. This decline corresponds to an increasing adoption of BI techniques (Fig. 4.2). The growing preference for BI approaches is likely due to their greater ability to model the complex, non-linear patterns present in EEG signals, thus improving the accuracy of MI classification.

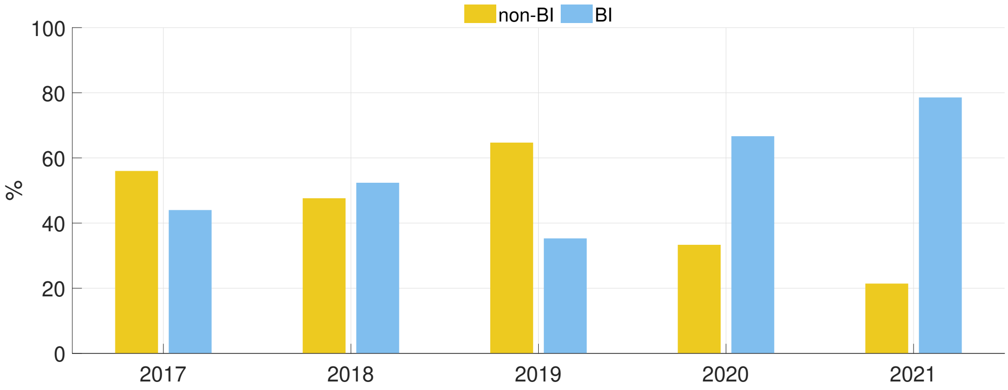


Figure 4.2: Percentage of non-brain-inspired (non-BI) and brain-inspired (BI) approaches proposed from 2017 to 2021.

4.2.3 Assessment methods and promising trends

The studies reviewed in this section reported their outcomes primarily in terms of classification accuracy or related metrics such as error rate and Cohen's kappa coefficient. It's important to clarify that classification accuracy refers to the proportion of correctly predicted target trials out of the total number of trials. In contrast, the error rate is the proportion of incorrectly classified trials. Meanwhile, Cohen's kappa coefficient adjusts for the likelihood of agreement occurring by chance, providing a more nuanced measure of classification performance [162]. It is important to distinguish classification accuracy from measurement accuracy, as defined in metrology. Measurement accuracy refers to the "closeness of agreement between a measured quantity value and a true quantity value of a measurand" [122].

Different strategies were used to evaluate classification accuracy, primarily involving the partitioning of datasets into training and evaluation sets. The most common methods for this purpose are:

- "*cross-validation*": the dataset is divided into "k" subsets. The model is trained on "k-1" subsets, and the remaining subset is used for testing. This process is repeated "k" times, with each subset being used once for validation. The final performance is reported as the average accuracy over all iterations, often together with a standard deviation to indicate the variability in the results.
- "*hold-out*": it involves a single split of the dataset into two parts. One is used for training the model and the other is used for evaluation. The classification accuracy is calculated once, based on this single partitioning.

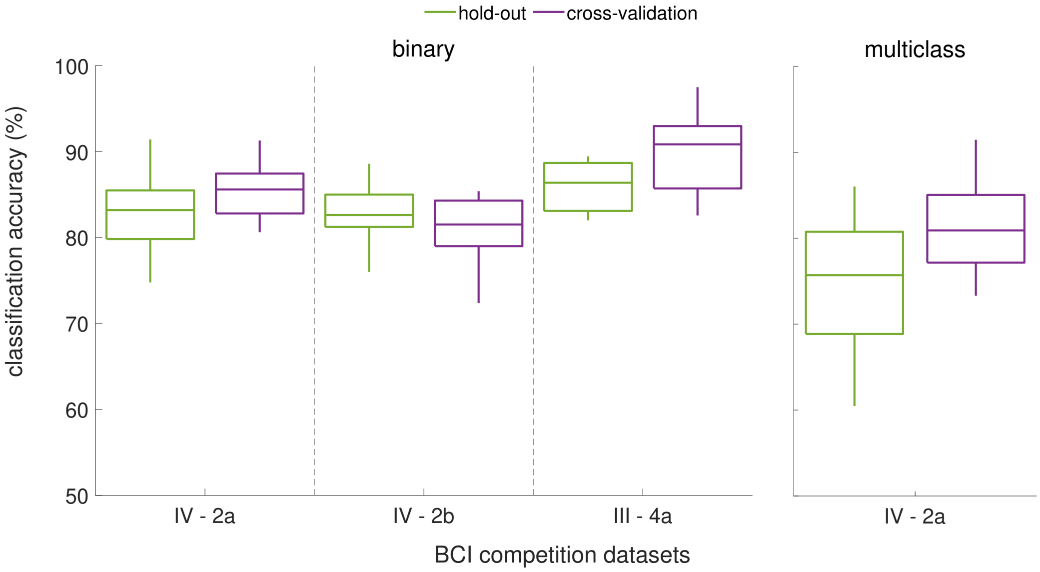


Figure 4.3: Mean classification accuracy reported by the reviewed studies. For the most frequently used datasets, results obtained with hold-out (green) and cross-validation (purple) are reported. Both the binary (on the left) and multiclass (on the right) case are presented.

According to the current review, 45 % of the studies employed cross-validation, with the number of folds varying between 5 and 30. The 37 % of the studies used the hold-out method. The remaining studies (8 %) either did not specify the method used or employed alternative methods (e.g. cross-validation on segmented trials or averaging results over several training and test splits). Additionally, 10 % of the studies did not provide details about their assessment method. A comparative analysis of these assessment methods was conducted, as illustrated in the box-plots shown in Fig. 4.3. These box-plots depict the distribution of reported mean accuracies across participants for each dataset. The analysis focused on the three most frequently used public datasets: "BCI Competition IV - Dataset 2a", "BCI Competition IV - Dataset 2b", and "BCI Competition III - Dataset 4a". This choice was made to maximise the number of studies available for comparison. It is worth noting that the first two datasets in Fig. 4.3 both pertain to "BCI Competition IV - Dataset 2a", but for different classification tasks: one for the classification of all four motor imageries included in the dataset and the other for the classification of left and right hand motor imageries.

Overall, the results show that cross-validation yields higher accuracy than the hold-out method, except in the case of the "BCI Competition IV - Dataset 2b". For this particular dataset, the data were more diverse, with EEG signals being recorded under different conditions, such as with and without feedback. This may account for the variation in performance outcomes.

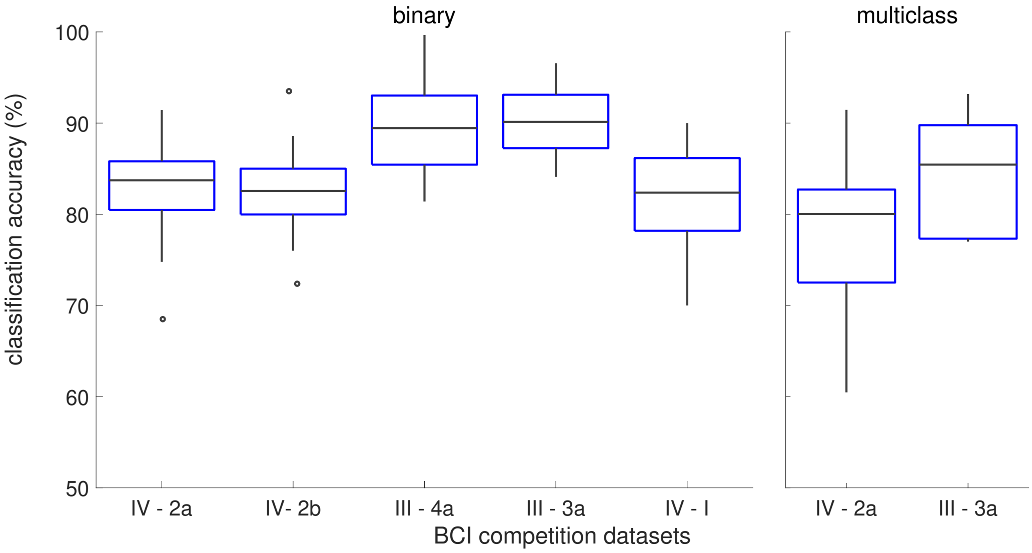


Figure 4.4: Mean classification accuracy reported by the examined studies for each of the most used datasets. Both the binary (on the left) and multiclass (on the right) case are presented.

In order to identify the most effective processing strategies for MI classification, datasets used in at least 10 % of the reviewed studies were considered. The datasets in question are “BCI Competition IV - Dataset 2a”, “BCI Competition IV - Dataset 2b”, “BCI competition III - Dataset 4a”, “BCI competition III - Dataset 3a”, and “BCI Competition IV - Dataset 1”. It is noteworthy that both the BCI competition IV-2a and BCI competition III-3a datasets include four MI tasks. Two studies were excluded from this analysis because they solely employed the “PhysionetMI dataset”, which did not meet the 10 % usage threshold [163, 164].

For each selected dataset, box-plots were made based on the reported mean classification accuracies across participants to visually compare performance distributions (Fig. 4.4). For binary classification, all possible pairs of classes were considered, while for multiclass classification, all available classes were considered.

The analysis focused on processing approaches above the 75th percentile (i.e., upper quartile), as these represent the most effective classification approaches with the highest accuracy rates. Additionally, particular attention was paid to the upper outliers of the box plots, which indicate levels of performance significantly above the general trend. In the present case, one such outlier was observed among the binary classifications when the “BCI Competition IV - Dataset 2b” was used.

Fig. 4.5 also provides insights into the chronological trends, processing strategies and evaluation methods used for the binary and multiclass classification scenarios. In the case of binary classification, the highest results were predominantly observed in the most recent publications, particularly in 2021. This suggests a trend of improved performance over time. Non-BI approaches dominated the studies with the best performance. However, it is important to note that despite their growing popularity, relatively fewer studies

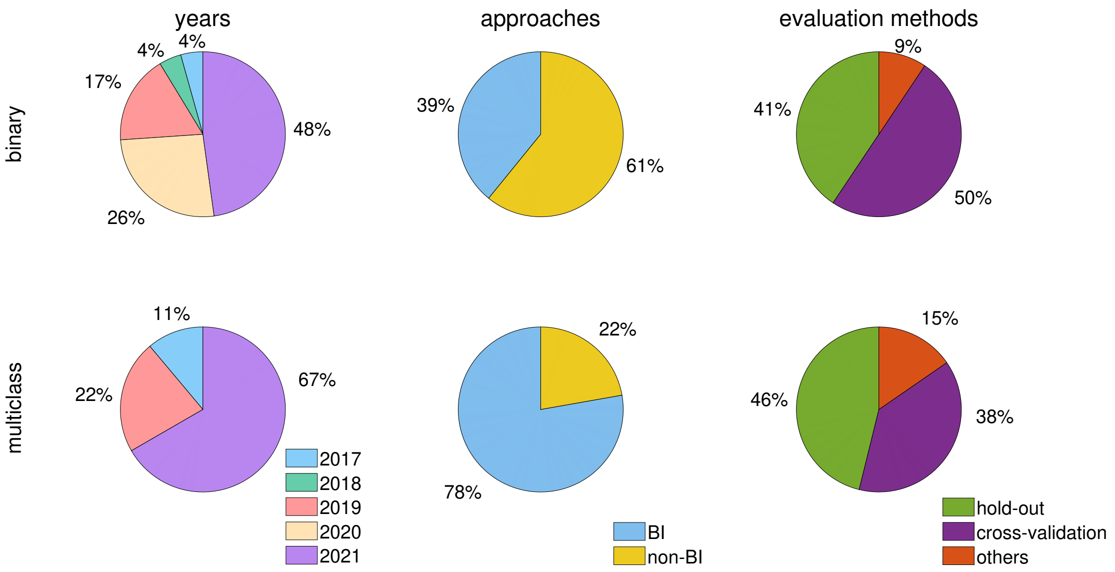


Figure 4.5: Analysis of most performing approaches.

have used BI approaches in the last five years. In terms of performance evaluation methods, cross-validation has been the most widely used method, reflecting its robustness in evaluating model generalisation across different data splits. In the case of multiclass classification, the highest results were similarly concentrated in recent years, again with a peak in 2021. This trend matches an increased focus on multiclass problems in the MI classification literature in recent years. Contrary to the binary case, BI approaches were more prevalent among the top performers for multiclass classification tasks. Additionally, the hold-out method was more frequently employed for performance evaluation in multiclass scenarios. This preference for hold-out over cross-validation is likely because cross-validation is typically used during the training phase of BI models to fine-tune hyperparameters and optimise network architecture. In contrast, hold-out is often preferred for the final performance testing of the trained network.

4.2.4 Successful strategies for motor imagery classification

A detailed analysis of the top-performing methods identified in the literature was conducted to identify the most effective processing strategies for MI classification. These approaches achieved classification accuracies within the 85 % to 100 % range for binary classification tasks and 83 % to 93 % for multiclass tasks. Using a type-A statistical method, the uncertainty associated with these performance metrics was calculated to be up to 6 % for binary case and up to 5 % for multiclass case. The type-A uncertainty was calculated by dividing the standard deviation of the results obtained from the papers under review by the square root of the number of participants included in the exploited dataset. A lower uncertainty level is indicative of a more homogeneous performance across subjects. It is noteworthy that the findings suggest no significant discrepancy in uncertainty between the most effective methods, indicating a compatibility between overall performance. Indeed, the ranges delineated with the mean accuracy and the associated uncertainty overlap. In terms of repeatability, the results varied by up to 7 % for binary tasks and 8 % for multiclass tasks. These percentage values were calculated by taking the difference between the highest and lowest classification accuracies reported by the top-performing approaches for each dataset. These values represent the upper tails of the box-plots depicted in Fig. 4.4, where the left subplot illustrates binary classification cases and the right subplot represents multi-class scenarios. Notably, the BCI Competition III-3a dataset exhibited the highest repeatability, with variability limited to 3 % for both binary and multiclass cases. However, this dataset's performance is based on a small sample size, raising concerns about the statistical significance of these results. Regarding overall variation in classification accuracy across different datasets, a reproducibility of 15 % was observed for binary tasks and 10 % for multiclass tasks. However, these values primarily refer to the most frequently used datasets, particularly those from the BCI competitions. The limited number of studies using different datasets, especially non-public ones, restricts further analysis. Additionally, missing information about the experimental setups in non-public datasets complicates direct comparisons, as noted in Tab. 4.1.

Following these initial observations, the analysis was further refined by separating the non-BI and BI approaches. The best-performing non-BI and BI strategies are summarised in Tab. 4.2 and Tab. 4.3, respectively.

The classification accuracies for non-BI methods fall within the 85 % to 100 % range for binary tasks and within the 86 % to 93 % range for multi-class tasks. Approximately 80 % of these studies employed feature extraction techniques based on CSP [165], which were sometimes combined with additional feature extraction techniques. Approximately half of the studies in question combined CSP with SVM classifiers [166]. These findings indicate that the combination of CSP and SVM remains a particularly effective approach, and that the most successful non-BI studies can be distinguished by their pre-processing and feature selection techniques. Notable pre-processing strategies include temporal segmentation of trials, a-priori channel selection over the sensorimotor cortex, and bandpass filtering, often across multiple frequency bands [151]. Additionally, the wavelet transform was employed frequently in these studies [167, 168]. Moreover, some studies that employed CSP and SVM also used the "least absolute shrinkage and selection operator" (LASSO) technique to refine feature selection by identifying the most discriminative features [169]. Notably, only two of the top-performing non-BI studies were applied to multi-class classification tasks. This could imply a limitation of non-BI approaches in handling more complex classification scenarios involving multiple classes.

For BI methods, the reported accuracies fall within the 82 % to 99 % range for binary tasks and within the 82 % to 95 % range for multi-class tasks. A single study [156] included in the analysis employed SNNs, achieving a maximum accuracy of 91 % in a multi-class scenario with an associated uncertainty of 4 %. The remaining studies employed ANNs, with approximately 70 % using DNNs. In particular, DNNs predominantly comprised convolutional layers. Additionally, other architectures were investigated, including "long short-term memory" (LSTM) networks [170], which are a type of "recurrent neural networks" (RNNs), and adversarial-based models [171]. In multiclass classification tasks, the mean accuracy for the most effective DNNs was 85 % with an uncertainty of 4 %. Non-DNNs did not rank among the top performers for multi-class tasks but showed a mean classification accuracy of 93 % with a 3 % uncertainty in binary tasks. This should be evaluated in comparison to the mean classification accuracy of 88 ± 4 % reported for DNNs. Although classification accuracy appears relatively consistent across different BI approaches when uncertainties are considered, there is a noticeable difference in mean accuracy between them. Further specific insights into repeatability and reproducibility are limited due to incomplete reporting for each sub-type of BI approaches. This topic is further discussed later in this chapter.

Pre-processing	Feature extraction	Feature selection	Classification	Dataset	
Normalization	Principal component analysis and cross-covariance	Best-first algorithm and correlation-based variable selection	Multi-layer perceptron	BCIc.III-4a (**B**)	[172]
A-priori time windows selection, BP and FB in [8,30] Hz	CSP	LASSO	SVM	BCIc.IV-1, BCIc.III-3a (**B**), BCIc.III-4a	[173]
Overlapped time windows, dual tree complex wavelet transform, in [4,8] Hz, [8,16] Hz and [16,32] Hz	CSP	Regularized neighbourhood component analysis	SVM	BCIc.IV-2a (**B**), BCIc.IV-2b, BCIc.III-3a (**B**)	[167]
BP in [8,30] Hz	CSP	Random Forest	Ensemble of random subspace k-NN	BCIc.IV-2a (**M**), BCIc.III-3a (**M**), BCIc.III-4a (**B**)	[174]
Channel and time window selection, FB in [8,26] Hz and BP in [8,24] Hz	Geodesic filtering CSP	Mutual information and Pearson correlation coefficient	SVM	BCIc.III-4a (**B**)	[175]
C3 and C4 a-priori selection, 1 s overlapping windows, BP in [8,14] Hz, [14,27] Hz and [27,45] Hz	Temporal absolute variations and segment-based bispectrum sums		SVM	BCIc.IV-2b, BCIc.III-4a (**B**), Non-public	[139]
Time domain staging and discrete Fourier transform-based decomposition.	Complex CSP	Energy and power-based selection, CSP sorting, Gaussian weighting, within-class scatter matrix, eigenvalues regularization.	Mahalanobis distance	BCIc.IV-2a (**B**), BCIc.IV-2b (**B**), BCIc.III-3a (**B**)	[176]
Time windows selection and BP in [8,30] Hz	CSP	LASSO	SVM	BCIc.IV-1 (**B**), BCIc.III-3a, BCIc.III-4a	[177]
Down-sampling, channel selection, BP in [8,30] Hz, multivariate wavelet transform, arrangement in vertical and horizontal sub-bands	CSP		LDA	BCIc.IV-1 (**B**), GS	[154]
Overlapping time windows selection, FB in [4,40] Hz	CSP and multiple-instance logistic regression	LASSO	SVM	BCIc.IV-2a (**B**)	[178]
BP in [4,40] Hz	Local activities estimation and regularized CSP	Power spectral density in [8,30] Hz with 4 Hz resolution	SVM	BCIc.IV-1 (**B**), BCIc.IV-2a, BCIc.III-4a, GS	[179]
15 constant-Q filters	CSP plus MAV, variance, and RMS in time domain, plus AR and PSR	ReliefF, Minimum redundancy maximum relevance, Fisher's method	SVM	BCIc.IV-2b (**B**)	[180]
C3, Cz and C4 a-priori selection, 2.0 s time windows with 1.9 s overlap, energy spectrum by wavelet packet decomposition and CDSF, FB in [4,40] Hz	CSP, STDF		Ensemble SVM	BCIc.IV-2a, BCIc.IV-2b (**B**)	[168]
C3 and C4 a-priori selection, trial segmentation	STDF, wavelet packet decomposition		SVM	BCIc.IV-2a, BCIc.IV-2b (**B**)	[181]
Overlapping time windows selection, BP in [8,40] Hz	CSP and AR	Kullback-Leibler divergence or Mutual information	NBPW	BCIc.IV-2a (**M**), BCIc.IV-2b	[182]

Table 4.2: "Most performing processing strategies using non-BI approach. **B** and **M** indicate the dataset for which the proposal has exceeded the 75th percentile of accuracies distribution for the binary and multiclass cases, respectively (see Fig. 4.4). AR: auto-regressive coefficient, BP: band-pass, CDSF: class discrepancy guided sub-band filter, CSP: common spatial pattern, FB: filter bank, k-NN: k-nearest neighbors, LASSO: least absolute shrinkage and selection operator, LDA: linear discriminant analysis, MAV: mean absolute value, PSR: power spectral ratio, RMS: root mean square, STDF: spatio-temporal discrepancy feature, SVM: support vector machine, BCIc.: BCI competition, GS: Giga Science" [16].

Pre-processing	Feature extraction	Feature selection	Classification	Dataset	
2-D continuous wavelet transform	CNN pre-trained by AlexNet on ImageNet dataset			BCIc.III-4a **(B)**	[155]
Multi-scale principal component analysis	Successive decomposition index		Feed-forward NN	BCIc.III-4a **(B)**	[183]
BP in [8,30] Hz	CSP	Particle swarm optimization and extreme learning machine		BCIc.IV-2a **(B)**, BCIc.III-3a **(B)**	[184]
Channel selection, FB in [8,32] Hz and in [6,30] Hz	CSP and LDA	Sequential backward floating selection	Radial basis function NN	BCIc.III-4a **(B)**, Non-public	[142]
Time window selection, BP in [8,30] Hz	Artifact rejected binary CSP	Self-regulated supervised Gaussian fuzzy adaptive system Art		BCIc.IV-2a **(B)**	[185]
Data augmentation	Electroencephalography-inception network (CNN based)			BCIc.IV-2b **(B)**, BCIc.IV-2a **(M)**	[186]
Data augmentation with a circular translation strategy	Feature separation network based on adversarial learning			BCIc.IV-2b **(B)**, BCIc.IV-2a **(M)**	[171]
FastICA, BP in [4,50] Hz, data augmentation with LGAN	Wavelet and fast Fourier Transform	Multi-output CNN and attention network		BCIc.IV-2b* **(B)**, BCIc.IV-2a **(M)**	[146]
Time window selection, BP in [0,40] Hz, signal normalization, Morlet wavelet transform	Multiscale space-time-frequency feature-guided multitask learning CNN			BCIc.IV-2b **(B)**, BCIc.IV-2a **(M)**, HG	[187]
Data augmentation with a circular translation strategy	CNN framework based on the discriminative feature learning strategy			BCIc.IV-2b **(B)**, BCIc.IV-2a	[188]
A-priori channel selection, common average referencing, subject-specific time window selection, FB in [7,30] Hz	CSP	F-score	Spiking NN	BCIc.IV-2a, BCIc.III-3a **(M)**	[156]
FB in [4,40] Hz	Dipole source estimation, time of interest selection, coordinate transformation, construction of 4D matrix		Three-dimensional CNN	BCIc.IV-2a **(M)**, ULM	[189]
FB in 43 bands	CSP	CNN		BCIc.IV-2a **(M),**	
μ and β bands	Multiple frequency CNN			BCIc.III-3a	[190]
Overlapping time window selection, FB in [4,38] Hz	CSP	CNN plus long short-term memory		BCIc.IV-2a **(M)**	[170]

Table 4.3: "Most performing processing strategies using BI approach. **B** and **M** indicate the dataset for which the proposal has exceeded the 75^{th} percentile of accuracies distribution for the binary and multi-class cases, respectively (see Fig. 4.4). * outlier in terms of classification performance. BP: band-pass, CNN: convolutional neural network, CSP: common spatial pattern, FB: filter bank, ICA: independent component analysis, LDA: linear discriminant analysis, LGAN: long short-term memory generative adversarial network, NN: neural network. BCIc.: BCI competition, HG: high gamma, ULM: upper limb movements" [16].

Overall, BI approaches have demonstrated superior performance in multi-class classification compared to non-BI approaches. Typically, deep learning models are trained directly on raw EEG signals [171, 186] or transformed data representations such as 2D or 3D arrays obtained from Fourier or wavelet transforms [146, 155, 187]. With a few exceptions [170, 190], feature extraction was performed manually for non-deep learning approaches. As with non-BI approaches, CSP or “Filter Bank Common Spatial Patterns” (FBCSP) were the most frequently used feature extraction techniques. Pre-processing strategies such as channel selection, temporal segmentation, and band-pass filtering were widely adopted for BI approaches as well. “In addition, unlike non-BI approaches, the need for a larger amount of training data emerged. Half of the best proposals using convolutional networks exploited data augmentation strategies. All of these were deep approaches. The used strategies were: (i) circular translation strategy [171, 188], where samples are circularly shifted by a fixed time step, (ii) the extraction of noise from a trial and its application to another trial [186], and (iii) long short-term memory generative adversarial networks (LGANs), which use random noise and label as input to generated realistic MI data [146]. Transfer learning was adopted as well in [155]. This approach consisted of a CNN architecture (AlexNet) pre-trained on the ImageNet dataset. The idea was to transfer previous knowledge on image classification to feature extraction and classification of EEG. Indeed, the final CNN is fine-tuned on an EEG dataset, but with a limited number of training samples” [16].

Finally, it is important to highlight a notable outlier within the analyzed studies. The study [146] demonstrated exceptional performance when using the BCI Competition IV-2b dataset. “In there, the authors proposed the usage of an ICA, a bandpass filter in the 4-50 Hz range, features extraction with wavelet and fast Fourier transforms, a multi-output convolutional neural-network, and an attention network for classifying the MI tasks. The proposal was implemented either with and without the LGAN data augmentation strategy” [16]. The authors achieved a classification accuracy of 93 % and 88 % with and without data augmentation, respectively. Notably, they are both above the 75th percentile for performance. The use of data augmentation techniques, specifically, appeared to be a critical factor contributing to its superior performance, positioning this proposal as one of the most effective BI approaches identified in this review.

4.2.5 Online experiments with self-collected data

As highlighted at the beginning of this section, only a limited number of studies evaluated their proposed methodologies using self-collected data. Among these, only two studies [138, 140] conducted online experiments to evaluate their algorithms under more realistic conditions. In [138], both synchronous and self-paced control modes were implemented to guide a robot according to MI classification results. In the synchronous control mode, the robot performed a movement only when the output of the classification corresponded to the desired task. During these synchronous experiments, participants were able to complete an average of 17 instructions within 41 s. Conversely, the self-paced control mode allowed the robot to execute a movement only after receiving two consecutive identical classification results. This mode introduces a delay to improve decision reliability, reducing the likelihood of errors due to sporadic misclassifications. In this case, participants managed to reach predefined goals in an average of 37 s. However, the study did not report specific accuracy metrics for the self-paced experiments. In [140], three participants used MI to control the turning of a wheelchair. The mean classification accuracy achieved was equal to 84 %. This study focused on a practical application scenario in which MI was used to assist with navigation tasks. Although they achieved a good level of accuracy on self-collected data, these two studies proposed methods that did not rank among the best performing approaches identified in this analysis, even when tested on public data. However, online experiments offers valuable insights into the practical usability of MI-based systems. Testing with self-collected data and in real environments is crucial for understanding how these systems would perform outside of controlled laboratory settings. Such evaluations help identify potential issues related to robustness, response time, and user adaptability, which are not fully captured in offline analyses.

4.3 Need for standardisation

The studies included in this analysis were reviewed with a focus on the state-of-the-art approaches and the essential components of successful processing strategies for MI classification in BCIs. The goal of this discussion is to consolidate existing knowledge and advancements in MI processing and provide guidance for developing effective MI-BCI systems. However, the analysis revealed several limitations due to variability in experimental setups and protocols, such as different types of MI tasks, variations in the number and positioning of EEG channels, and differences in the number of participants. These factors contribute to challenges in achieving repeatability and reproducibility of the results across studies. This section will discuss the key achievements of the current analysis, while also highlighting areas where further research is needed to address unresolved issues and challenges.

4.3.1 Key achievements

The investigation of effective strategies for EEG processing was conducted in two main phases. The initial stage of the analysis involved an examination of the various approaches and assessment methods employed in the included studies. Subsequently, the most effective approaches were identified by focusing on widely used public datasets in order to ensure a statistically robust comparison. This required a sufficient number of comparable studies to be available for analysis. In this context, 87 out of 89 studies were considered in the second step, as they exploited at least one of the frequently used datasets. The remaining two studies were excluded from the analysis as they were based solely on the PhysioNet EEG dataset [163, 164], which had too few proposals to allow for a statistical analysis. It is noteworthy that the studies in question achieved performance levels comparable to the median performance on the BCI Competition IV-2a dataset when classifying right-hand versus left-hand imagery in a binary manner. However, a dataset-specific analysis for PhysioNet would be beneficial as the classification accuracy can be heavily influenced by the number of available participants.

It is noteworthy that studies focusing on binary classification tasks were distributed relatively evenly over the five-year period covered by this review. In contrast, the majority of studies on multiclass classification were published in 2020 and 2021. In both the binary and multiclass classification domains, the analysis revealed that the most successful approaches emerged primarily in the most recent years under consideration. Specifically, five studies from 2021 [171, 186, 146, 187, 174] were identified as exceeding the 75th percentile in terms of performance for both binary and multi-class classifications. Of these, four studies employed BI approaches. Additionally, two other recent studies [176, 184] demonstrated high performance across all datasets for binary classification. It is also important to mention that the approaches described in [184, 185] were among the top-performing for binary classification across different MI tasks. However, their performance in multi-class cases fell below the median. It is not feasible to conduct a more comprehensive analysis at this stage, due to the limited number of proposals that were evaluated on several public datasets or with self-collected data.

The findings indicate that both BI and non-BI approaches can achieve compatible classification accuracy. However, there is a growing interest in BI approaches in recent years. In the context of non-BI approaches, the existing literature consistently identifies CSP as a frequently employed feature extraction technique [10]. The present analysis further corroborates the association between CSP and high classification performance, particularly when combined with LASSO for feature selection and SVM for classification. These methods benefit from leveraging existing knowledge about neurophysiological processes, which makes non-BI approaches particularly effective for binary classifications. However, these methods have limitations, including the need for meticulous feature extraction and selection. The main advantages of non-BI approaches are their simplicity in implementation, reduced training time, fewer parameters to optimise, and reliability even with small datasets.

On the other hand, BI approaches represent a newer trend in the field and typically do not require extensive prior knowledge in signal processing. Within BI methodologies, two main sub-fields are emerging: SNN and ANN. SNN are gaining attention because they closely mimic the operational principles of the human brain. Their advantages include continuous real-time computation capabilities and the ability to integrate multiple dimensions of information (e.g., temporal, spatial, frequency, and phase) into a single model, as well as the capacity to handle large volumes of data. Furthermore, they are well-suited for hardware implementation and online applications [191]. Although their potential as a comprehensive processing pipeline has been recognised, only one of the top-performing studies [156] employed a SNN. Other high-performing BI approaches have predominantly used ANN. Similar to non-BI approaches, non-deep ANN in BI approaches are limited to classification of binary tasks. However, they are relatively easy to implement due to their simpler architecture, which involves few layers and parameters. Finally, it is becoming increasingly common to employ deep neural networks, which comprise multiple hidden layers. Despite the inherent complexity of these architectures, their primary advantage lies in their ability to extract and classify features autonomously from the data [192].

The results of this analysis demonstrate that deep BI approaches are more effective than other strategies in multi-class MI classification tasks, while achieving results that are comparable to non-BI approaches in the classification of binary tasks. Deep learning methods offer several advantages, including the ability to process raw data directly with minimal preprocessing and to undertake all stages of the process, from feature extraction to classification, within a single integrated pipeline. However, these approaches require extensive training due to the large number of parameters involved, necessitating substantial computational resources and potentially limiting their applicability for online applications. Additionally, deep learning models require large datasets, which are not always readily available in the BCI domain. To address the challenge of limited training data, the Tab. 4.3 shows that two main strategies were employed: "data augmentation" and "transfer learning". Data augmentation techniques have proven effective in enhancing model robustness and mitigating overfitting issues associated with small datasets. Transfer learning, on the other hand, leverages knowledge acquired from previous learning tasks to reduce training time and computational costs. This strategy is particularly promising for MI-based BCIs, as it can help minimise the calibration of the system each time a user has to use it. Overall, BI approaches are very promising for multiclass classification tasks, which represent one of the most important challenges in the field of MI-BCI [48]. As illustrated in Fig. 4.5, most of top-performing proposals for multiclass cases belong to the BI category. In particular, SNN and deep learning strategies are recommended for enhancing the number of commands in MI-BCIs. Notably, SNN appear to be particularly suitable for online applications due to their speed and flexibility. Therefore, while non-BI approaches remain well-established for binary classification tasks, BI approaches offer a promising avenue for addressing the more complex challenges of MI-BCIs, such as the need to accurately classify multiple commands in real-time.

Researchers are addressed to explore the processing pipelines detailed in Tab. 4.2 and Tab. 4.3 to achieve successful EEG classification in MI-BCI applications. “They usually report many details on the EEG data acquisition, and it was surely possible to rigorously quantify classification performance in terms of accuracy, uncertainty, and repeatability of the results. However, as already mentioned, current studies often lack reproducibility because experimental conditions are not always completely reported” [16]. Additionally, most research has predominantly used datasets from BCI competitions. To advance the field, a broader range of datasets should be employed, given that MI-related EEG data is heavily influenced by factors such as the training of subjects. A more comprehensive validation that includes a wider sample of subjects would provide a more robust assessment of proposed processing strategies. Ideally, researchers should evaluate their methods not only on benchmark datasets, such as those from BCI competitions, but also on other public datasets and their own self-collected experimental data, particularly from online experiments. This approach would facilitate a direct comparison with existing state-of-the-art techniques while also allowing for a deeper exploration of the limitations of different methods by analysing how performance varies under different conditions. Tab. 4.1 outlines the main factors influencing MI-related EEG acquisition, and it could serve as a reference for future experiments that aim to control experimental settings. Furthermore, there is a pressing need for standardisation of performance assessment methods. The literature reveals a variety of assessment methods, including various cross-validation strategies and differing data split methods in hold-out approaches. This variability hampers direct comparison of results across studies. However, it is encouraging to note that recent studies are increasingly adopting a dual approach: employing cross-validation for model definition (e.g., optimising hyperparameters) and validation, as well as testing on independent data using a hold-out strategy. Although challenges remain, particularly due to the non-stationarity of EEG signals, this dual approach enables a more nuanced understanding of MI processing. Regarding the standardization efforts, the BCI research community has long recognized the importance of establishing common paradigms to drive technological progress. One notable initiative is the “Mother of All BCI Benchmarks” project, which proposes an open-source software suite as a standardized framework incorporating benchmark data and state-of-the-art processing algorithms [193]. Such initiatives could significantly facilitate comparisons across different studies and promote the development of new algorithms. Despite their potential benefits, these frameworks have not yet been widely adopted in the field. Another critical point to consider is that MI-BCI performance is often reported almost exclusively in terms of classification accuracy. The associated standard deviation or uncertainty metrics are frequently not explicitly reported, which limits the understanding of the variability of the results. Furthermore, other metrics such as those assessing speed and latency are commonly used in evaluating other BCI systems but are rarely applied to MI-BCIs. This gap can be partly explained by the fact that MI detection generally requires longer time windows for data acquisition and processing compared to other BCI paradigms [7]. Nonetheless, metrics like the “Information Transfer Rate” (ITR) could provide valuable insights into the efficiency of MI-BCIs, as has been demonstrated in other types of BCI systems [48]. Reducing target detection time is a key objective for enhancing the ITR in BCI systems [6]. Therefore, it would be worthwhile to explore MI-related phenomena further to identify control applications that could benefit from these improvements.

Chapter 5

Measurement of user engagement

Brain-computer interfaces (BCIs) were originally used as a means to assist people with motor disabilities or paralysis. Thanks to a BCI, these individuals were able to communicate with the outside world and indirectly recover motor functions.

However, the inherent direct access to brain activity provided by BCIs presents a broader opportunity to improve human-computer interaction. Rather than supplant existing interaction techniques designed for direct and explicit control, BCIs offer an additional implicit communication channel.

In the context of motor imagery-based BCI, it is important to monitor user's level of engagement. This is closely related to the actual generation of motor imagery-related brain activity. Especially in the rehabilitation field, a decline in patient engagement during rehabilitation sessions based on motor imagery can compromise the effectiveness of the therapy. Measuring the user's level of engagement becomes important to identify optimal intervention points, such as introducing breaks in the rehabilitation session or providing additional support to strengthen engagement levels and improve the overall effectiveness of the system.

Some of the work described in this chapter has been previously published in [23].

5.1 Rationale for measuring user engagement

In the context of human-computer interaction, engagement refers to a positive behavioural, cognitive, and emotional state experienced by users during their interaction with a system or an application [194]. It comprises various dimensions, including workload, attention, motivation, interest, and emotional response [195, 196].

Measuring user engagement is interesting in several contexts. For example, in the field of education, understanding student engagement can help educators tailor their teaching methods to better meet the needs and interests of their students, leading to improved learning outcomes [194, 197, 198]. Another example is in the context of entertainment. A video game might introduce obstacles or increase the speed of the game depending on the user's level of engagement in order to keep it high [199]. Measuring engagement is also important for evaluating the effectiveness of interventions and treatments, particularly in healthcare and rehabilitation settings. For example, in rehabilitation therapy, maintaining high levels of patient engagement is essential to achieving positive outcomes. By monitoring engagement levels during therapy sessions, clinicians can identify when patients are becoming fatigued or disinterested and adjust treatment accordingly to maximise effectiveness [200].

5.1.1 Engagement assessment methods

Traditionally, engagement has been assessed through evaluation grids, self-assessment questionnaires, interviews, and direct observation, whether in field settings or controlled environments [201]. Although widely used, these traditional methods have limitations. For instance, self-report questionnaires allow users to provide subjective ratings of their engagement levels based on their perceptions and experiences. While self-report measures have traditionally been the most widely used methods, they are subjective. Direct observations provide insights in real time. However, they can be invasive and can alter participants' behaviour leading to biased results.

In recent years, there has been a great deal of research into biosignal-based measurements, such as galvanic skin response, heart rate, and electroencephalogram. They provide an objective way of assessing user engagement by directly measuring physiological responses associated with cognitive and emotional states [196, 202, 203, 204, 205]. In particular, electroencephalography (EEG) has emerged as a promising tool for measuring engagement, as a wide range of information on the subject's mental state can be derived from it. The Engagement Index (EI) is a well-known parameter in the literature that provides a quantitative measure of user engagement. It is derived from EEG signals and it is defined as:

$$EI = \frac{\beta}{\alpha + \theta} \tag{5.1}$$

where β is the beta-band power (13 Hz to 30 Hz), α is the alpha-band power(8 Hz to 13 Hz) and θ is the theta-band power (4 Hz to 8 Hz) [196, 204, 205, 206, 207].

This index is based on the assumption that increased beta-band activity is associated with attentive mental states, such as those involved in cognitive tasks, visual system activation or movement planning [196]. Conversely, an increase in alpha and theta activity is associated with a decrease in vigilance and mental alertness. Alpha-band decreases often indicate information processing, while increases indicate a resting state [208].

5.1.2 Eliciting engagement

The evaluation of user engagement typically requires the administration of particular tasks or activities intended to elicit responses related to engagement. While there are no specific tests designed solely for the purpose of eliciting engagement, several existing assessments are commonly used at this aim. These tests often focus on cognitive performance, attentional processes and response inhibition, which are closely related to the engagement.

One example is the Continuous Performance Task (CPT) [209]. It is a widely used assessment tool for evaluating sustained attention and vigilance. By presenting a series of stimuli over time, the CPT challenges individuals to maintain focus and respond to specific target stimuli while ignoring distractors. Variations of the CPT were adapted in a variety of fields, including psychology, neuroscience and educational research, to measure engagement-related constructs.

Another commonly used test is the Stroop test, which assesses cognitive flexibility and inhibitory control [210]. In this task, participants are asked to name the colour in which words are written, while the words themselves may be incongruent colour names (e.g. the word 'red' written in blue). Successful completion of the Stroop test requires individuals to inhibit automatic responses and selectively attend to relevant information, thereby engaging cognitive processes associated with attentional control.

The d2 test is a widely used measure of attention and concentration [211, 212]. It requires participants to quickly identify specific characters (e.g. the letter ‘d’ with two dashes) while ignoring similar distractors. The test measures processing speed, accuracy and the ability to sustain attention over time, making it relevant for assessing engagement-related constructs.

The N-back test is a cognitive task that evaluates working memory and attentional control [213]. Participants monitor a sequence of stimuli, such as letters or numbers, and indicate when the current stimulus matches the one presented ‘n’ steps back in the sequence. This task challenges working memory capacity and the ability to update and manipulate information in real-time, providing insights into cognitive engagement and executive functioning.

Some examples of these tests are shown in Fig. 5.1.

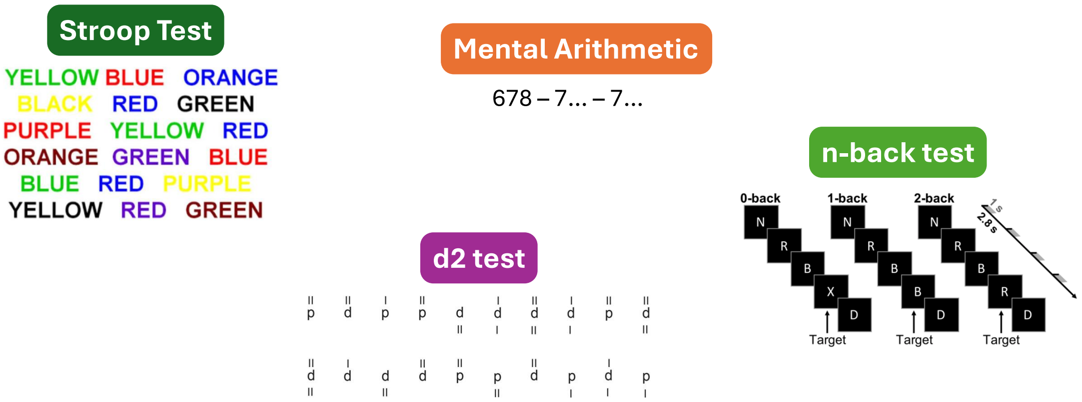


Figure 5.1: Examples of existing assessments commonly used to elicit engagement.

5.1.3 Challenges

Although the EI is widely used to measure engagement, it does not fully account for all dimensions of engagement [194, 214]. This limitation highlights the need for alternative methods that can provide a more comprehensive understanding of engagement dynamics.

Recent research has introduced innovative approaches to address the limitations of traditional engagement measures. For instance, in [194], cognitive and emotional engagement during learning was successfully detected using a filter bank Common Spatial Patterns (CSP) approach. This method resulted in better performance when compared to the results obtained using EI.

Similarly, in [214], a hybrid system that combines eye-tracking and engagement-based BCI technology was proposed for a hands-free version of the Tetris game. The system leverages a filter bank and a CSP to extract features, detecting user engagement levels and dynamically adjusting the speed of the game accordingly.

These advances highlight the potential of new methodologies to overcome the limitations of existing engagement measures.

Moreover, measuring engagement poses a challenge due to the lack of standardised stimuli designed to elicit engagement. While various cognitive tasks and assessments can indirectly evoke engagement responses, the development of targeted stimuli adapted to different contexts requires further exploration.

Finally, assessing engagement across different contexts remains a significant challenge.

5.2 Design and implementation of the engagement-based BCI

A passive engagement-based BCI is proposed. The aim is to measure user engagement through neural oscillations that are acquired by means of a wearable EEG instrument. A computerised version of the d2 test was adapted to increase user engagement levels and obtain distinguishable features from the EEG signal for training a machine learning algorithm. The system's generalisability and applicability in different contexts are subsequently assessed. In particular, the effectiveness of the trained algorithm in estimating user engagement levels in real time during various tasks is assessed.

5.2.1 Modified version of d2 test

The d2 test is a commonly used measure of selective attention and processing speed [212]. It was originally developed by Brickenkamp in 1981 to evaluate an individual's capacity to concentrate and differentiate between relevant and irrelevant stimuli within a limited time frame.

In the conventional d2 test, participants are presented with a sequence of characters, each containing the letter ‘d’ with two marks around or below it in any order. These are interspersed with distractor characters, the letter ‘p’. The task assigned to the participants is to rapidly scroll through the characters and indicate each instance of the letter ‘d’ with two dashes (the ‘d2’ pattern), ignoring any other characters or variations of the target pattern. The test is presented in the pen and paper version in its classical format.

To enhance engagement during the d2 test, a computerised version with modifications was developed. The goal was to make the task more challenging and interactive, thus increasing participants' motivation and engagement.

In the modified version, additional distractor characters were introduced. Specifically, the new distractors were the letters ‘b’ and ‘q’ with one to four marks above and below them. This increases the cognitive load and demands greater attentional control from participants. Furthermore, the proposed test comprises 12 lines with 57 items per line to be completed within one minute, intensifying the time pressure and requiring rapid decision-making. The computerised version provides users with a feedback on their performance. Correctly identified d2 patterns are shaded in blue, while incorrect selections are shaded in red. Finally, a progress bar in blue displays the remaining time for completing the task, encouraging participants to maintain focus and pace. However, it is worth noting that while this feature aims to enhance motivation and engagement, it may also induce stress due to the time constraint.

An example of the proposed d2 test is presented in Fig. 5.2.

Figure 5.2: Example of the proposed d2 test to elicit user engagement.

5.2.2 Experimental protocol

The experimental protocol was designed to elicit engagement and to assess it during different tasks. To elicit engagement, the d2 test paradigm was proposed, which alternated between engagement and resting states. This paradigm consisted of two sets of the modified version of the d2 test and two sets of fixation cross, each lasting one minute. They were presented in a randomised order. A short 5-second break separated each round to allow participants a brief relax. The system training phase lasts slightly over 4 minutes. To assess the trained system, the participants were presented with a new d2 test paradigm. It involved a new independent EEG recording. Then, participants were asked to play Tetris game and watch videos. The experimental procedure is shown in Fig. 5.3.

Prior to the experiments, participants received instructions on how to perform the d2 test and play Tetris, including examples. For the d2 test, participants were instructed to perform it line by line from left to right as quickly as possible to maintain high levels of engagement. During the fixation cross, they were instructed to relax as much as possible while keeping their eyes fixed on the cross.

At the end of the first d2 test paradigm, the collected EEG signals were used to estimate the classification accuracy of the proposed algorithm in distinguishing between two mental states, i.e. engagement and relaxation. A minimum classification accuracy threshold of 75 % was set to consider the acquired data as informative. If the threshold was not achieved, the first d2 test paradigm was repeated to obtain new data. A new classification accuracy was estimated based on the new data. If participants did not reach the minimum threshold a second time, the experiment continued anyway.

The d2 test paradigm was repeated for each participant to evaluate the algorithm's performance on new independent data. To assess the reproducibility of the system, the algorithm trained on data from the previous d2 test paradigm was employed to estimate participants' engagement levels during distinct tasks, such as playing Tetris and watching

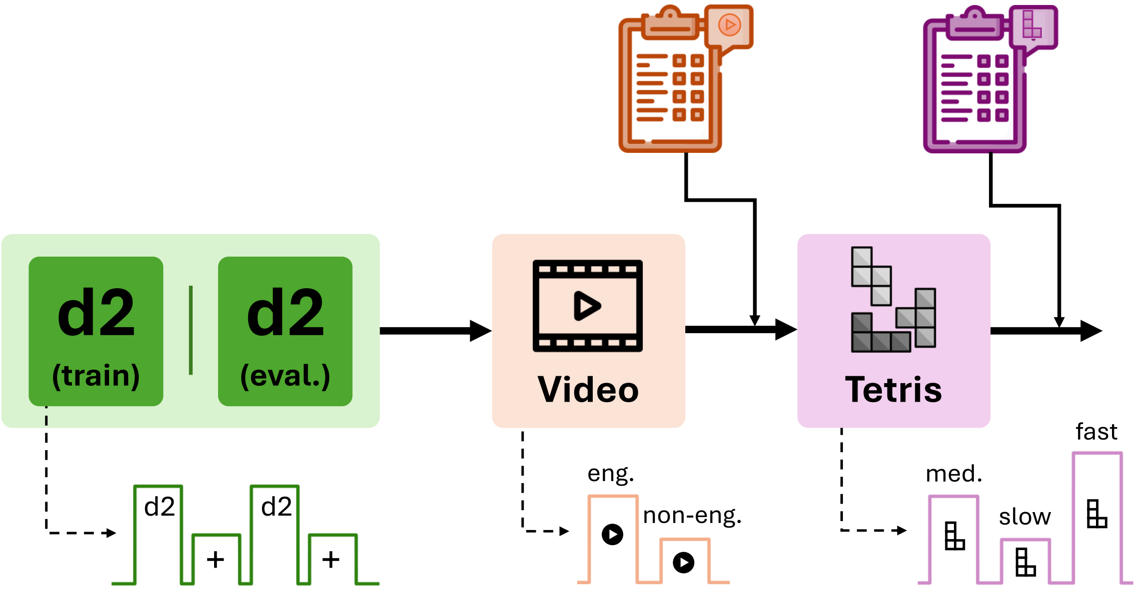


Figure 5.3: The experimental procedure comprised two recordings of the d2 test paradigm followed by tasks involving playing Tetris and watching videos. The d2 test paradigm alternated between engaging and resting states, using the d2 test and a fixation cross.

videos.

The Tetris game was played at three different speeds: slow, medium, and, fast. Each speed lasted for one minute and was followed by a 30-second pause. The order of presentation for the three speeds was randomised to avoid potential biases. For the video-watching task, participants viewed two one-minute videos without sound. They were selected to represent engaging and relaxing stimuli. The two videos were labelled differently based on their content. One was labelled as engaging because its scenes changed quickly and it was a short advertisement, while the other was labelled as relaxing due to its natural landscape scene. The order of video presentation was randomised and they were interspersed with a 5-minute break. No specific instructions were provided to participants during this phase. The Tetris and video paradigms were also presented in a randomised order.

After completing each Tetris and video paradigm, participants were asked to complete a subjective questionnaire. This was designed to assess their perceived engagement specifically in relation to the task they had just completed. The questionnaire assessed several dimensions of the definition of engagement [195, 196], using a 7-point Likert scale [215]. These dimensions included:

- the perceived challenge of the task;
- the subjective perception of the time taken to complete the task;
- the emotional involvement evoked by the task content;
- the ease of maintaining concentration throughout the task;
- the visual appeal of the task stimuli.

Participants provided separate ratings for each dimension corresponding to the three Tetris speeds (slow, medium and fast) and the two video stimuli (short advertisement and landscape). Notably, after viewing the two videos, participants were not asked the first question about perceived challenge.

5.2.3 Proposed system

EEG instrument

The EEG signal was measured by means of the wearable EEG-headset Unicorn Hybrid Black (g.tec medical engineering GmbH, Austria). It consists of eight input channels. EEG electrodes are hybrid, meaning they allow for both dry and wet EEG acquisition. In the case of wet acquisition, conductive gel is used to reduce electrical impedances between the electrodes and the skin. The EEG signal is acquired at a sampling frequency of 250 Hz and with 24 bit resolution. Data is transferred from the EEG instrument to a personal computer via Bluetooth.

According to the international 10–20 system [216], the standard electrode configuration involves placing the electrodes in Fz, C3, Cz, C4, Pz, PO7, Oz, and PO8. However, for this study, electrodes were placed in the parieto-occipital region of the scalp. Specifically, electrodes were placed at P3, P4, PO3, POz, PO4, O1, Oz, and O2 (Fig. 5.4). This decision was driven by two main reasons: i) the electrodes on motor area can be influenced by mouse movements during the d2 test; b) the parieto-occipital area is less influenced by eye movements and blinking than the fronto-central area. In addition, several studies have demonstrated the involvement of the parieto-occipital area in different cognitive processes (e.g., working memory, visual and sensory information processing, particularly in relation to arousal) [206, 217, 218, 219, 220].

Ground and reference were positioned on the subject's left and right mastoids, respectively, by means of disposable surface adhesive electrodes.

The EEG signal was acquired using wet electrodes due to challenges posed by achieving good electrode-skin contact in the densely hair-covered occipital scalp area.

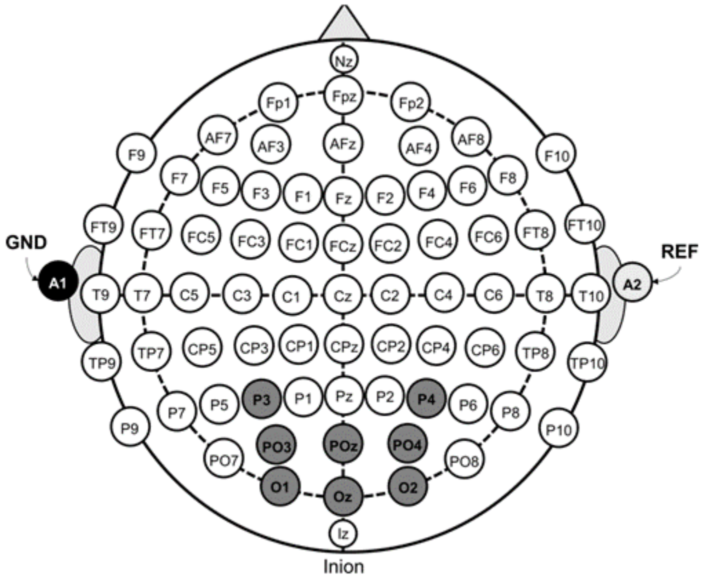


Figure 5.4: Electrode placement in the parieto-occipital region of the scalp.

Developed application

A software application was developed to manage the presentation of stimuli, acquire EEG signals, and process them in real-time. The application measures the user's engagement levels and provides real-time feedback on their engagement status. Before starting experiments, the signal quality can be assessed in the application by mounting the EEG helmet on the user. Ongoing information about noisy channels is also provided. The user is asked to select the paradigm to be carried out. As a first step, the d2 test paradigm should be chosen to train the proposed algorithm. The paradigm consists of an alternating d2 test and fixation cross.

After completing the d2 test, the application calculates several reliable performance metrics [212, 221, 222]. These include the *total processed items* (which is the sum of the number of items processed), the *error of omission* (which is the sum of the number of target items not cancelled), the *total correctly processed* items (which is the total number of items processed minus the total number of errors made), the *accuracy* (which is the total number of correctly processed items divided by all processed items), and the *concentration performance* (which is the total number of correctly canceled items minus the total number of incorrectly canceled items). In contrast to the traditional paper-and-pencil version, where scores are calculated for each of the 14 test lines [222], the performance scores in this study were calculated by considering all the processed items within one minute, from the first item until the last selected one.

For the fixation cross, the application displays a fixation cross at the center of the screen and tracks time progression using a blue bar, similar to the d2 test. Following the first d2 test paradigm, the application assesses the effectiveness of the classification model in distinguishing between engaging (d2 test) and resting (fixation cross) states. Additionally, the application provides the option to retrain the model if necessary.

The calibration process, which includes mounting the EEG instrument on the user, d2 test execution, and model training, typically takes about 5 minutes. Once trained, the model can assess user engagement in real-time during any activity, both within and outside the application.

Proposed algorithm

To enable online experiments, the EEG processing pipeline was optimised for computational efficiency while maintaining accuracy in estimating engagement levels. This was achieved through the adoption of a filter bank Common Spatial Patterns (CSP) approach for engagement estimation (Fig. 5.5).

The raw EEG recordings underwent initial pre-processing steps to enhance signal quality. To mitigate power line interference, a notch filter with a cutoff frequency of 50 Hz was applied using a 2nd order Butterworth filter. A bandpass filter bank covering the frequency ranges from 4 Hz to 8 Hz, from 6 Hz to 10 Hz, and from 8 Hz to 12 Hz was then applied to the data. It is important to note that higher frequency components were excluded from the analysis due to their susceptibility to contamination by muscle artefacts induced by participants' posture during the visually demanding d2 test.

The EEG data was then segmented into non-overlapping 1-second windows. This resulted in a resolution of 1 s for online analysis. The segmentation strategy produced 120 segments for each class, corresponding to the d2 test and resting states. The CSP algorithm was used to extract features from these EEG segments. It is a signal processing technique commonly used in the analysis of EEG signal. The CSP algorithm is based on the idea that unique patterns of brain activity are associated with different mental states or tasks. These patterns are reflected in the spatial distribution of EEG signal power across channels. The algorithm identifies spatial filters that increase the variance of EEG signals for one class (e.g. engaged state) while decreasing the variance for another class (e.g. resting state). This enhances the distinction between the two classes.

The extracted features were used to train a Linear Discriminant Analysis (LDA) classifier. LDA is a supervised machine learning technique commonly used for dimensionality reduction and classification tasks. It aims to identify a linear combination of features that can effectively distinguish between different classes in a dataset. The key principle underlying LDA is to maximise the separation between classes while minimising the variation within each class. It achieves this by projecting the original feature space onto a lower-dimensional subspace where class separability is optimised. LDA provides binary labels reflecting the predicted class and continuous numerical scores indicating class separability. In the current work, binary labels are used to assess the accuracy of the classifier, while continuous scores are used to estimate the level of subject engagement.

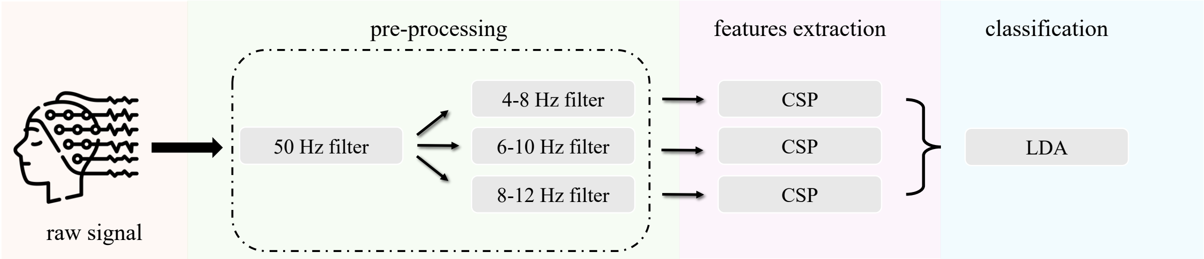


Figure 5.5: EEG processing pipeline based on a filter bank Common Spatial Patterns (CSP) approach for engagement estimation. LDA: Linear Discriminant Analysis.

As described in the previous section, following the initial d2-test paradigm, the application assesses the classification model's efficacy in distinguishing between engagement states (d2-test) and rest (fixation cross). The classification accuracy was estimated using the non-overlapping 1-second EEG segments of this first paradigm, employing the proposed algorithm in a 10-fold cross-validation technique. In this study, the model was retrained once if the cross-validation accuracy fell below 75 %.

After the training of the proposed algorithm based on the first d2 test recording, it was used online to estimate the subjects' engagement during the Tetris and Video recordings (Fig. 5.3). In this case, the features were extracted from non-overlapping 1 s windows and the LDA scores were used to quantify the estimated engagement.

Statistical analysis

"The statistical analyses were performed using MATLAB R2021b (MathWorks Inc., United States). The normality of data was tested using the Shapiro–Wilk test [223]. The statistical test was chosen according to the normality of the sample, so either the paired t-test or the Wilcoxon signed rank test was used.

Descriptive statistics are reported as mean and the standard deviation (SD), or median and the interquartile range (IQR) (i.e., 25th and 75th percentile). The Bonferroni correction was used in the case of multiple comparisons to control for type I errors.

The difference in subjects' d2 test performance between the two d2 test recordings was investigated. Therefore, the obtained performance scores were compared between the two recordings and the p-values were Bonferroni corrected. The perceived engagement was analyzed based on the sum of the engagement dimensions of the Tetris game questionnaire. Additionally, the scores of the classification model during the four conditions (rest period, low speed, medium speed, and fast speed) were compared in pairs recordings and the p-values were Bonferroni corrected. Perceived engagement during the two videos was also analyzed based on summing up engagement dimensions of the questionnaire for the two videos, as well as the scores of the classification model during the two conditions (engaging and non-engaging video)" [23].

5.2.4 Results

Participants

Twenty-three healthy individuals participated in the experimental study, comprising seven females and 16 males with an average age of 34 ± 7 years. All participants had normal or corrected vision. They did not received any prior training related to the experiment. All participants read and signed an informed consent form before taking part in the experimental study.

d2 test paradigm

All subjects achieved the estimated classification accuracy of 75 % after the first d2 test paradigm. Thus, it was not necessary to re-train the model by acquiring new data.

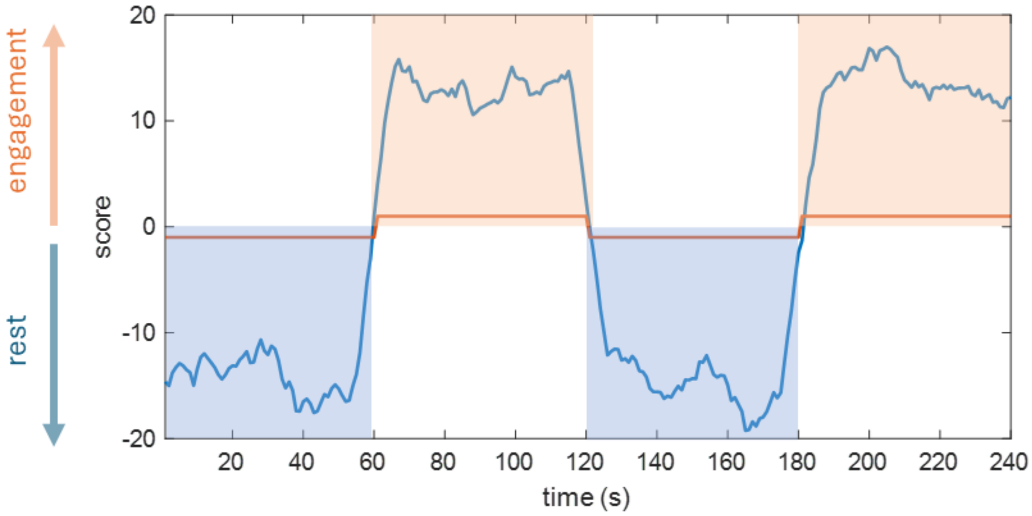


Figure 5.6: S01: Algorithm score in distinguishing between the d2 test and the fixation cross during the independent recording (i.e., the second d2 test paradigm). The blue curve shows the score, while the orange line indicates the task performed during the experiment.

Fig. 5.6 shows the results obtained for the subject S01 on an independent recording. The x-axis represents the time in seconds, while the y-axis represents the score of the algorithm. The blue curve shows the score, with negative values corresponding to the relax phase and positive values to the engagement phase.

The orange line indicates the task performed during the experiment, taking a value of -1 during the fixation cross (relax phase) and 1 during the d2 test (engagement phase).

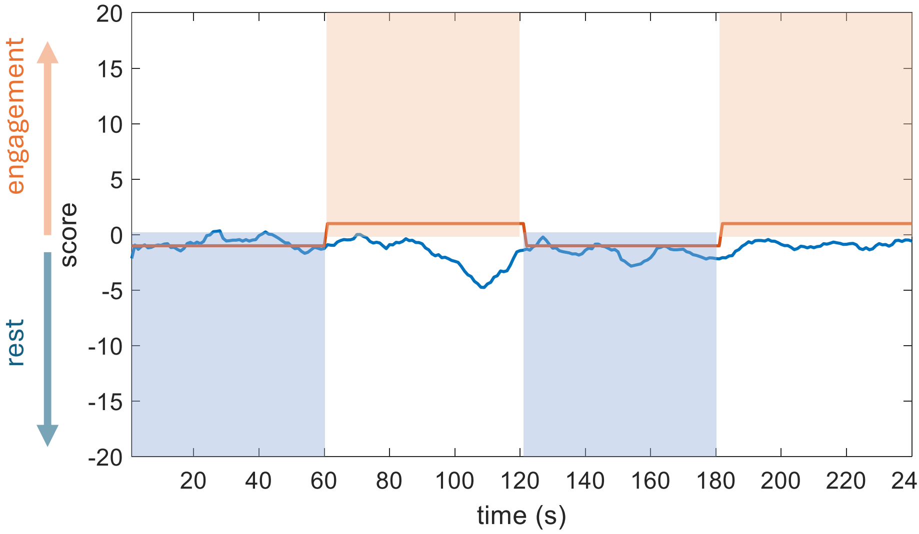


Figure 5.7: S07: Algorithm score in distinguishing between the d2 test and the fixation cross during the independent recording (i.e., the second d2 test paradigm). The blue curve shows the score, while the orange line indicates the task performed during the experiment.

The score curve has a noticeable transition between these two phases, which is reflected in the abrupt change of the curve. The rectangles in the figure, shaded in different colours, visually distinguish the relax and engagement phases. The light blue rectangle represents the relax phase, corresponding to the fixation cross task, and is characterised by negative scores indicating lower levels of engagement. In contrast, the light orange rectangle represents the engagement phase, corresponding to the d2 test task, with positive scores reflecting increased engagement. Furthermore, the constancy of the subject's level of engagement and relax during the respective phases can be observed. This consistency underlines the algorithm's ability to detect both the state of relax and the state of engagement, as well as the intensity of these mental states in the subjects.

For comparison purposes, Fig. 5.7 shows the results obtained for the subject S07 on an independent recording. In contrast to S01, S07 does not show any increase in engagement throughout the test and instead maintains a steady state of relaxation. The score curve remains consistently negative, which indicates that this subject doesn't move into the engagement phase.

Fig. 5.8 illustrates the classification accuracy in distinguishing between the d2 test and the fixation cross during the independent recording (i.e., the second d2 test paradigm). Additionally, Fig. 5.8 includes results obtained by incorporating an extra feature selection step post-feature extraction using CSP and using the engagement index. These last two analysis were conducted offline for comparative purposes. Both analysis were computed for the same non-overlapping 1 s windows and channels. Feature selection aimed to emphasize the most significant features, potentially enhancing classification performance by focusing on those contributing most to the model's discriminant function. Features with weight vector values above the 75th percentile or below the 25th percentile were retained, while those within the interquartile range were set to zero. Instead, the engagement index was performed according to Eq. 5.1. For all approaches, classification accuracy for each subject alongside the mean classification accuracy and their respective type A uncertainty

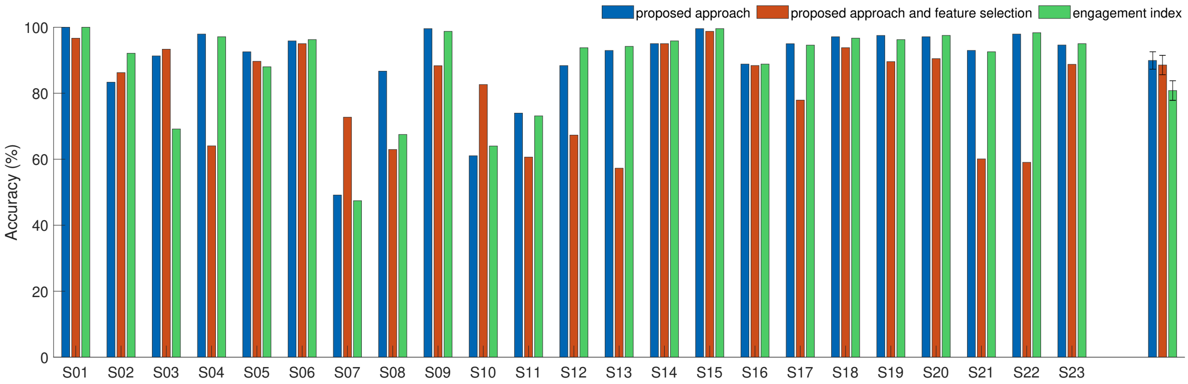


Figure 5.8: Classification accuracy in distinguishing between the d2 test and the fixation cross during the independent recording (i.e., the second d2 test paradigm). The results obtained using the proposed algorithm (in blue), the proposed algorithm with a further step for select the features (in red), and the approach based on the engagement index (in green) are reported. The error bars on the means reflect the type A uncertainty.

were reported.

The mean results ranged from 90 ± 3 % when the CSP-based approach was used to 81 ± 3 % when the offline engagement index was used. Out of the 23 subjects, 20 achieved over 80 % accuracy in independent registration using the proposed algorithm, while 14 achieved the same using the engagement index. The proposed approach has a mean classification accuracy that is approximately 10 % higher than when using the engagement index.

Introducing the feature selection step to the proposed approach resulted in a mean accuracy of 89 ± 3 %. Similar to the engagement index, 14 subjects achieved over 80 % accuracy in independent registration with this method. The performance with the proposed approach including feature selection is comparable to the results obtained without it. While this demonstrates the robustness of the feature selection method in maintaining high classification performance, the original proposed approach without feature selection is preferred for online processing due to its reduced computational complexity and fewer processing steps.

Measure	Training	Evaluation	Δ	p-value
Total processed items	466.6 (86.7)	456.2 (78.4)	-10.3 (53.4)	1.000
Omissions	6.0 (4.7)	4.1 (3.2)	-1.8 (4.3)	0.281
Total correctly processed	460.3 (85.1)	451.6 (77.1)	-8.7 (52.3)	1.000
Accuracy (%)	97.6 (1.7)	98.2 (1.3)	0.6 (1.7)	0.535
Concentration Performance	18.8 (7.5)	18.4 (7.7)	-0.3 (8.5)	1.000

Table 5.1: "d2 test performance scores computed for the training and evaluation recording. The results are reported across subjects as mean (SD). The difference in performance scores (Δ) is computed as training minus evaluation. Depending on the normality of the differences, the Wilcoxon signed rank test or paired t-test were used for statistical testing. Finally, the p-values were Bonferroni corrected" [23].

"In order to also take into account how subjects performed the d2 test during the two recordings, the performance scores calculated by the application were evaluated. Tab. 5.1

shows the results as mean (SD) for each recording and for the difference of them. No difference in d2 test performance between the two recordings was found" [23].

Tetris

Fig. 5.9 shows the engagement level of a representative subject across three speeds of the Tetris game. The score of the algorithm trained with the d2 paradigm is reported. The x-axis represents time in seconds, while the y-axis represents the algorithm's score.

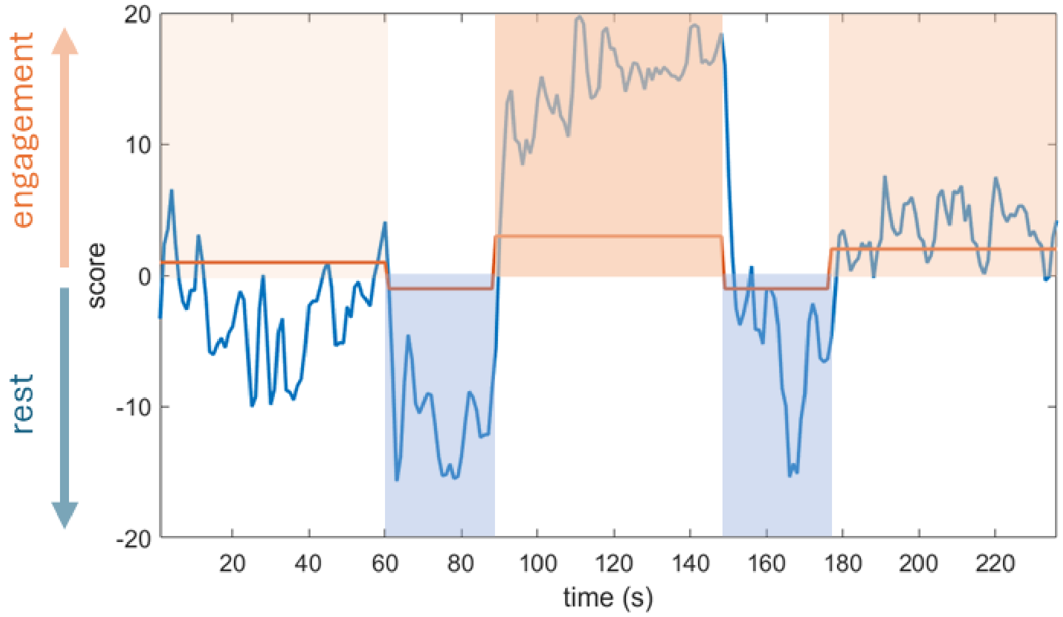


Figure 5.9: Algorithm score in distinguishing between three speeds of the Tetris game (i.e. slow, medium, and fast speed). The blue curve shows the score, while the orange line indicates the task performed during the experiment.

The orange line represents the task performed during the experiment. It takes a value of -1 during the 30-second pauses (relax phase) and 1, 2, and 3 during the slow, medium and fast speeds of the Tetris game, respectively. The blue curve illustrates the score, with higher values indicating greater engagement.

It is noteworthy that during the slow speed, the algorithm suggests that the subject's engagement was low, which is consistent with the expected boredom due to the slow pace of the game. Conversely, the subject's engagement remained consistently high throughout the medium and fast speeds, indicating active participation and interest. Interestingly, the algorithm detects a higher level of engagement during the fast speed compared to the medium speed, highlighting the challenge posed by the increased pace of the game.

Finally, a 30-second interval between the different speeds resulted in a relaxation of the subjects, as evidenced by the negative scores observed during this interval.

The rectangles in the figure, shaded in different colours, serve to distinguish the phases of relaxation and engagement. The colours of the rectangles relating to the three different speeds of Tetris exhibit an increasing intensity as the speed increases.

Fig. 5.10 presents the subjects' responses to the questionnaire (7-point Likert) on perceived engagement. It depicts the median values of the 23 subjects for each of the five dimensions of engagement considered and for each of the three speeds of Tetris.

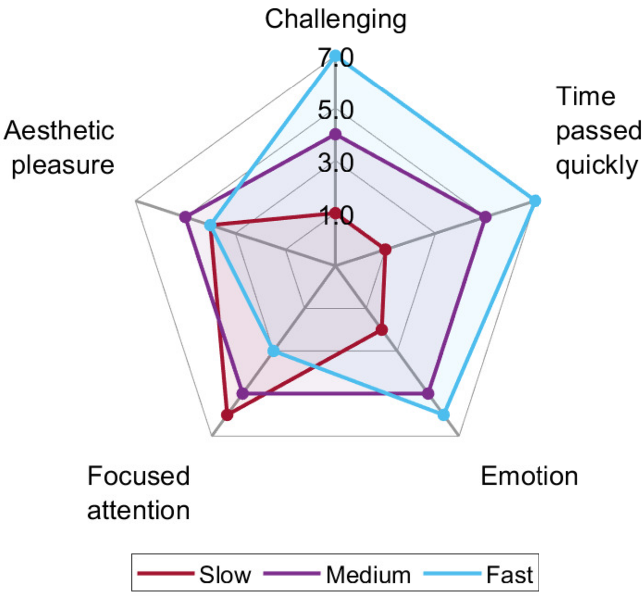


Figure 5.10: Subjective results across subjects for the Tetris paradigm. Responses from the questionnaire (i.e. perceived engagement) are reported as the median value for each dimension of engagement and for each speed (i.e. slow, medium and fast).

"The sums of the engagement questionnaire dimensions were observed to be 14.5 (3.2), 22.9 (3.8) and 24.9 (4.7) points in mean (SD) for the slow, medium, and fast speed, respectively. Subjects experienced the medium and fast speeds as more engaging than the slow speed ($p < 0.001$, respectively), whereas no difference was observed between medium and fast speeds ($p = 0.081$)" [23].

Fig. 5.11 presents the objective results of the subjects' level of engagement while playing Tetris. These results were estimated by the proposed algorithm as the LDA score for each condition. The x-axis represents the different game conditions, while the y-axis represents the classification model scores. Negative scores are associated with resting states, while positive scores are associated with engaging states. The boxplots show the median scores for each subject. Overall, there is a distinction between the different conditions, with predominantly negative scores during the pauses between the three speeds. On the contrary, the median scores for the three speeds of Tetris are mostly positive.

Tab. 5.2 shows the results of the statistical analysis comparing the difference between the four conditions in terms of subjects' estimated engagement in pairs. "As expected, the three Tetris speeds were found to be more engaging than the rest period ($p < 0.001$, respectively). The same applies to the medium and fast speeds compared to the slow speed ($p < 0.001$, respectively). Similarly, as for the subjective questionnaire, no difference was observed between the medium and fast speeds ($p = 0.345$). Taken together, subjects' perceived engagement is in line with the objective evidence put forward by the classification model scores" [23].

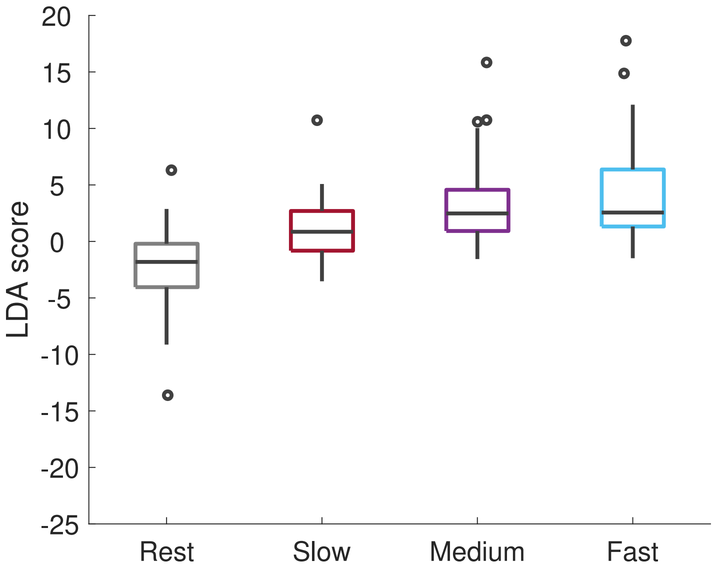


Figure 5.11: Objective results across subjects for the Tetris paradigm. The subjects' engagement was estimated by the proposed algorithm. Specifically, each boxplot reflects the median LDA scores of the subjects for each condition of the game.

Tetris	Δ	p-value
Rest < Slow	2.2 [0.4, 5.5]	< 0.001
Rest < Medium	3.6 [2.1, 6.7]	< 0.001
Rest < Fast	4.2 [2.1, 7.1]	< 0.001
Slow < Medium	1.5 [0.7, 3.8]	< 0.001
Slow < Fast	2.0 [0.6, 3.3]	< 0.001
Medium = Fast	0.1 [-0.5, 0.8]	0.345

Table 5.2: "Tetris paradigm: Across subject statistical analysis using the Wilcoxon signed rank test. Subjects' estimated engagement (median LDA score) obtained by the proposed classification model were used. The difference in estimated engagement (Δ) for the respective comparison was calculated by subtracting the latter condition from the former (e.g., Slow minus Rest). The differenes are reported as median [IQR]. Depending on the normality of the differences, the Wilcoxon signed rank test or paired t-test were used for statistical testing. Finally, the p-values were Bonferroni corrected" [23].

Video

Fig. 5.12 presents the results obtained from a representative subject. The x-axis denotes time in seconds, while the y-axis represents the algorithm score.

The orange line, corresponding to the task performed during the experiment, takes a value of -1 during the landscape video and 1 during the short movie. The blue curve depicts the score, where negative values indicate the relax phase and positive values indicate the engagement phase. Additionally, two rectangles, shaded in different colours, are used to visually delineate the relaxation and engagement phases. The same algorithm, trained on the d2 test, is employed to measure the subjects' engagement levels while viewing two different videos. The landscape video elicits initial engagement from the

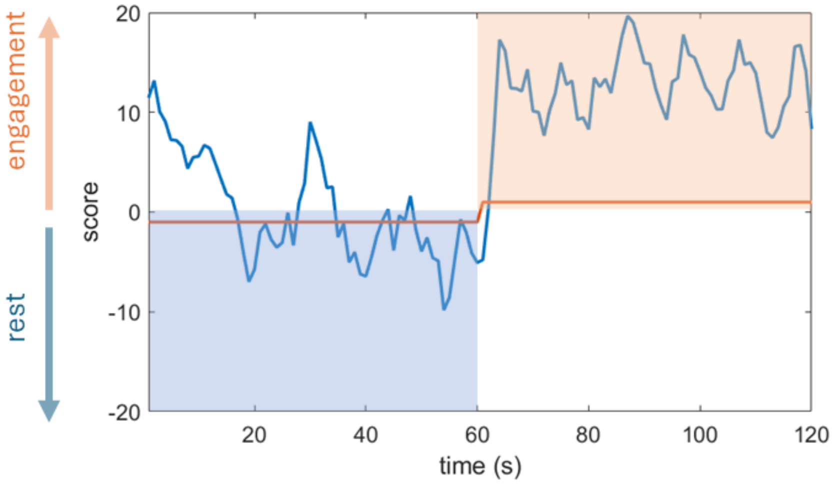


Figure 5.12: Algorithm score in distinguishing between two videos (i.e. engaging and non-engaging). The blue curve shows the score, while the orange line indicates the task performed during the experiment.

subject, as evidenced by the spike in engagement at the beginning. This aligns with the expectation that novel stimuli capture attention. Subsequently, engagement gradually decreases, likely due to the repetitive nature of the video. Conversely, engagement remains consistently positive throughout the duration of the short movie, indicating sustained involvement.

Fig. 5.13 presents the subjects' responses to the questionnaire (7-point Likert) on perceived engagement. It depicts the median values of the 23 subjects for each of the four dimensions of engagement considered in this case and for each of the two videos. The "Challenging" dimension is not reported in this case as it was not a question in the video paradigm. "The sum of the engagement questionnaire dimensions during the video paradigm was observed to be 16.7 (4.6) and 20.3 (5.1) points in mean (SD) for the landscape and advertisement videos, respectively. Subjects experienced the advertisement video to be overall more engaging than the landscape video ($p = 0.025$)" [23].

Fig. 5.14 presents the objective results of the subjects' level of engagement while watching the two videos. These results were estimated by the proposed algorithm as the LDA score for each condition.

The x-axis represents the different videos, while the y-axis represents the classification model scores. As for the Tetris paradigm, negative scores are associated with resting states, while positive scores are associated with engaging states. The boxplots contains the median scores for each subject. Overall, the median scores are predominantly negative or close to 0 during the non-engaging video. On the contrary, the median scores for the advertisement video are mostly positive.

"Finally, the across subject delta resulted in 1.8 [1.2, 3.5] in median [IQR] comparing the difference between the advertising and landscape videos. Consistent with the subjects' perceived engagement, the advertisement was observed to be more engaging by the classification model ($p < 0.001$)" [23].

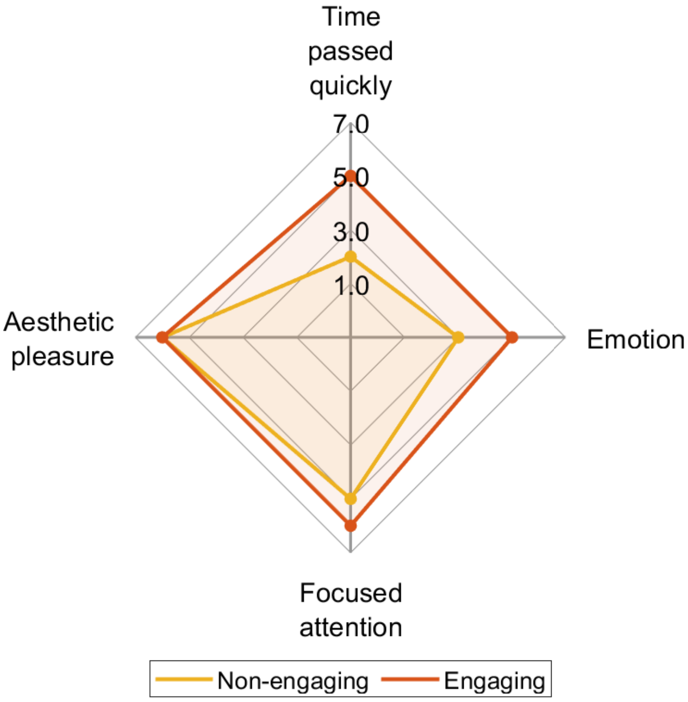


Figure 5.13: Subjective results across subjects for the Video paradigm. Responses from the questionnaire (i.e. perceived engagement) are reported as the median value for each dimension of engagement and for each video (i.e. landscape and advertisement videos). The "Challenging" dimension is not reported in this case as it was not a question in the video paradigm.

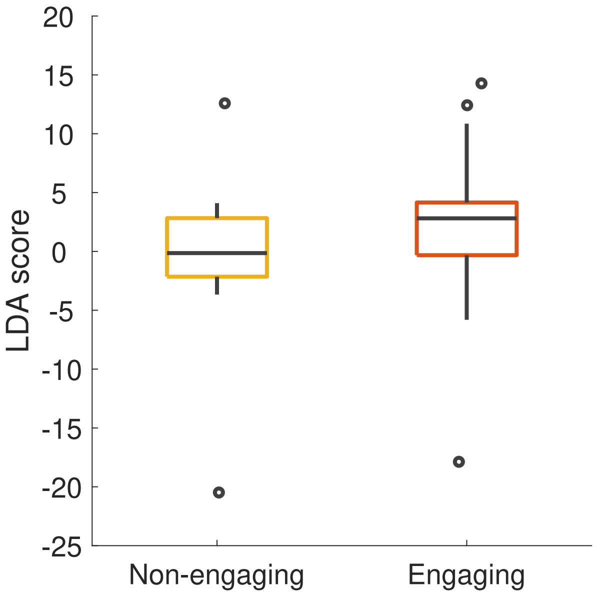


Figure 5.14: Objective results across subjects for the Video paradigm. The subjects' engagement was estimated by the proposed algorithm. Specifically, each boxplot reflects the median LDA scores of the subjects for each video.

5.2.5 Discussion

This study demonstrates the capability of the proposed methodology to differentiate between states of engagement and rest. In accordance with existing literature [194], it also affirms that the classification accuracy of the engagement index is lower to that of the filter-bank CSP algorithm. Notably, the proposed approach yields approximately 10 % higher accuracy across subjects compared to the engagement index. This discrepancy in classification accuracy could be attributed to two factors: (i) the beta band exhibited smaller modulations over parieto-occipital regions compared to more frontal regions [224], and/or (ii) a positive correlation between theta power and engagement was observed in this study.

The scores derived from the d2 test performance (see tab.5.1) indicate that subjects processed an average of approximately 4 lines per minute with low omission errors and high accuracy. The concentration performance index remained consistent throughout the two consecutive recordings.

Although standard performance metrics were employed [212, 221, 222], direct comparison with previous literature is challenging due to the utilisation of a modified version of the d2 test in this study. It is worth remarking that the proposed d2 test duration is shorter (2 min) than the original version (4 min). However, the scores were calculated based on both d2 tests administered in each recording. Consequently, the total number of items processed, the number of omissions and the total number of items processed correctly remain comparable to the results in the literature. [212, 221]. Finally, it is not possible to make a direct comparison of the concentration performance score due to the enhanced complexity of the proposed d2 test, which includes more distractors than the original version. Consequently, the number of correctly deleted items in the proposed version is significantly lower than that reported in the literature for the original d2 test. Furthermore, there were no discernible discrepancies in the performance scores of the d2 test between the two recording sessions, as the tests were administered in close succession.

The previously trained algorithm was employed to measure subjects' engagement levels during Tetris gameplay and while watching two different types of videos. Understanding engagement in gaming is of pivotal importance for enhancing the player experience and holds significance in serious gaming applications, such as learning facilitation and rehabilitation interventions. Similarly, measuring engagement while watching videos is promising for predicting population preferences, particularly in neuromarketing activities.

This study found that perceived engagement and engagement estimated by the algorithm differed depending on the speed of Tetris and video stimuli. Among the speeds at which the game Tetris was played, the slow pace was perceived as boring due to its slowness. However, it allowed meticulous positioning of the blocks. In contrast, the fast speed posed a challenge, as subjects had difficulty manipulating the blocks. Both video stimuli were perceived as engaging. This is potentially attributable to the subjects' novelty in seeing them, which prompted them to pay more attention to the details of the scene. It is likely that engagement with the landscape video decreased over time due to the repetitive nature of the content, whereas the rapid scene changes in the advertising video maintained engagement throughout.

Subjective responses to the questionnaire regarding perceived levels of challenge and perception of time were consistent with these results (Fig. 5.13 and 5.10). Subjects indicated that emotional responses were greater when playing Tetris at medium or fast speed, or when watching advertising video, than when playing Tetris at slow speed or watching landscape video. Notably, no significant difference in aesthetic enjoyment was found between the two videos and the three speeds of Tetris, contrary to previous research [225]. This approach could be useful in adapting applications according to users' estimated levels of engagement, as proposed in [214]. However, it is important to note that no quantitative results were presented in the cited work.

Chapter 6

Wearable neural interface for self-paced motor imagery gaming

6.1 Rationale for developing a self-paced system

6.1.1 Synchronous versus asynchronous systems

Motor imagery-based brain-computer interfaces (BCIs) allow users to control devices by imagining specific movements. For effective operation, the BCI system needs to process brain signals in real-time, provide accurate feedback, and function without fixed time constraints. However, many BCIs proposed in literature are used offline and depend on synchronous paradigms [226, 227, 228].

In a synchronous BCI, users are instructed to imagine specific movements of body parts within set time frames [16]. The processing of brain signals is constrained to these predetermined or variable time intervals rather than the entire acquired signal. The mental task is predefined and linked to an external cue or trigger. For this reason, this paradigm is also known as “cue-based”. It simplifies the processing and classification challenge since brain signals are collected under controlled conditions. This makes easier to differentiate between different types of motor imagery. Nevertheless, synchronous systems constrain the user, requiring them to adapt to fixed intervals to express their intentions [10]. This can reduce autonomy and may feel unnatural, particularly in real-world applications.

On the other hand, asynchronous BCI offer greater flexibility by allowing users to imagine movements at any time, without depending on external cues [16]. They are also known as “self-paced BCI”. Once the system is trained, it is designed for continuous real-time processing of brain signal [229]. By removing temporal constraints, self-paced BCIs improve user autonomy and provide a more natural interface. They are more suitable for real-world or extended reality situations where interactions happen unpredictably and continuously. In rehabilitation, self-paced systems are well-suited for ongoing tasks like controlling wheelchairs or exoskeletons [230]. They are also effective for applications that require cursor control on a screen or controlling an avatar [231, 232]. Despite the aforementioned benefits, developing robust, accurate, and reliable self-paced BCIs is still a major challenge. This is due to the complexities involved in continuous brain signal processing and the need to reduce false detections while maintaining responsiveness.

6.1.2 Challenges

BCIs based on motor imagery represent a field of research that is still evolving. These devices have the potential to make a significant contribution to medical research, particularly in the field of neurorehabilitation. Events such as the Cybathlon have become a focal point for the development of self-paced MI-BCI systems. It provides an opportunity for researchers to evaluate their solutions in realistic settings and for patients to use state-of-the-art assistive technologies [233]. These competitions also allow for the exploration of the effects of long-term training on potential end-users, a topic that has not been extensively covered in existing research [234, 235].

A significant challenge in this kind of competitions is the prohibition on using voluntary artifacts, like intentional eye blinks or muscle contractions, to control the system. Although this rule ensures that BCIs depend solely on neural activity, it also highlights the limitations of current MI-BCIs in achieving reliable control. Many systems proposed in the literature take advantage of such artifacts to improve reliability. In [232], the voluntary blinking of the eyes was used for error correction. Additionally, in [236], three consecutive blinks were used to turn the online processing of brain signals on or off. The authors also implemented a buffer to minimize the risk of false positives.

A cue-based asynchronous BCI to induce cortical plasticity in stroke patients was proposed in [237]. In their approach, patients imagined dorsiflexing their paretic ankle while receiving electrical stimulation as feedback. However, the system temporarily disabled the classifier after detecting a motor imagery task, limiting its asynchronous functionality and increasing response time. Similarly, the authors in [238] also pause their classification processes temporarily when errors are detected, such as a collision in a virtual environment. Other systems used hierarchical classifiers for self-paced control [231]. Their method involved using an initial classifier to identify “no movement” before engaging a second classifier to differentiate between specific tasks, such as stepping left or right. These approaches aimed to reduce false positives but ultimately delayed user interaction and diminished the overall responsiveness of the system.

Hybrid BCIs, which merge motor imagery with additional signals or paradigms, represent another approach to overcoming the limitations of MI-BCIs [239]. These systems frequently combine EEG with other biosignals, such as functional near-infrared spectroscopy [240] or electromyography [241], or integrate motor imagery with paradigms like steady-state visual evoked potentials [242] or P300 potentials [243]. Although these methods can improve accuracy and broaden the range of commands, they also increase system complexity, which can pose challenges for implementation and user training. The hardware used in MI-BCIs also plays a crucial role in their effectiveness and practicality. Most systems employ 10 to 30 electrodes to improve spatial resolution [236, 235]. However, while these configurations enhance signal quality, they are often impractical for daily use due to the lengthy setup times and reduced portability. Efforts to simplify hardware configurations must balance maintaining performance with achieving a practical design for end-users.

Finally, previous studies rarely took into account detailed data on the online use of the system. This limits an in-depth evaluation of performance.

6.2 Proposal

The previous chapters of this thesis lay the groundwork for the design choices made in this proposal. The goal is to create a virtual reality endless runner game featuring two levels of increasing difficulty. The game is controlled through a self-paced MI-BCI. The control commands rely solely on the user's ability to imagine movements of body parts or to relax, which enhances the naturalness and ease of interaction. Additionally, the proposed system is designed to be highly wearable. An EEG instrument with only eight electrodes primarily placed over the sensorimotor area was used. This setup is beneficial due to its reduced complexity and minimized user discomfort, making the system more practical for real-world applications. Finally, to ensure the usability of the data collected during gameplay, all game events were synchronized with EEG signals and recorded for thorough online performance evaluation. Examples of game events include the timing and position of targets, the player's trajectory, and whether targets were collected or missed.

Based on the analyses conducted in Chapter 2, Device 1 was identified as the most appropriate EEG system for this application. It demonstrated the best performance among the evaluated instruments, with a mean delay of around 20 ms and an associated uncertainty of about 2 ms. These features make Device 1 well-suited for real-time applications, ensuring accurate synchronization between the EEG signals and the game events.

Chapter 3 focused on techniques for artifact removal and examined the challenge of minimizing artifacts in low-density EEG setups while maintaining the underlying brain activity. A new algorithm was proposed that effectively removes artifacts even with a limited number of EEG channels. However, it is only applicable in offline scenarios due to the processing delay it introduces. This limitation underscores the existing gap in achieving artifact-free EEG signals in real-time. Incorporating artifact removal techniques into an online processing pipeline inevitably introduces some delay. Initially, the signal has to be acquired before it can be corrected, and the processing time for artifact removal techniques can vary from a few milliseconds to over 70 ms. To tackle this issue, alternative strategies such as optimal electrode placement and user training to reduce movement were implemented to lessen artifact contamination during online operation.

Chapter 4 reviewed the current state of motor imagery-based brain-computer interface signal processing. It demonstrated that filter bank and common spatial pattern (FBCSP) algorithm remain the most effective for binary classification tasks. Instead, convolutional neural networks (CNNs) show promise for multiclass problems. However, few of these algorithms have been validated for online applications. In particular, CNNs require extensive training due to their high number of parameters, leading to significant computational demands that hinder their practicality for real-time use. Additionally, the variability in their performance makes them less reliable for consistent operation. As a result, classical methods like FBCSP were selected for this system because of their established reliability and efficiency in online applications.

Finally, Chapter 5 presented a new BCI system designed to evaluate user engagement. This system provides important insights into tracking how engaged users are while playing a VR endless runner game. Monitoring this engagement could lead to useful feedback for adapting the system, ultimately enhancing the user experience and performance over time.

6.2.1 Experimental protocol

The experimental protocol consisted of two main phases conducted on the same day: a calibration phase and a game phase. The entire experiment lasted about 55 min. Before the experiment began, participants received verbal instructions detailing the procedures, the purpose of the experiment, and the concept of motor imagery. They were asked to perform different hand movements, such as squeezing, playing the piano, dribbling a basketball, or lifting weights. Participants were encouraged to choose the movement they felt most confident imagining and to focus on the kinesthetic sensations associated with it. Kinesthetic imagery has been found to engage the sensorimotor cortex more directly than visual or abstract imagery [244]. Once they chose a specific movement, they were instructed to maintain their concentration on that movement throughout the experiment.

Following this preparation, participants were fitted with the EEG helmet by an experienced operator. The helmet was not removed until the end of the experiment. A schematic overview of the experimental session is shown in Fig. 6.1.

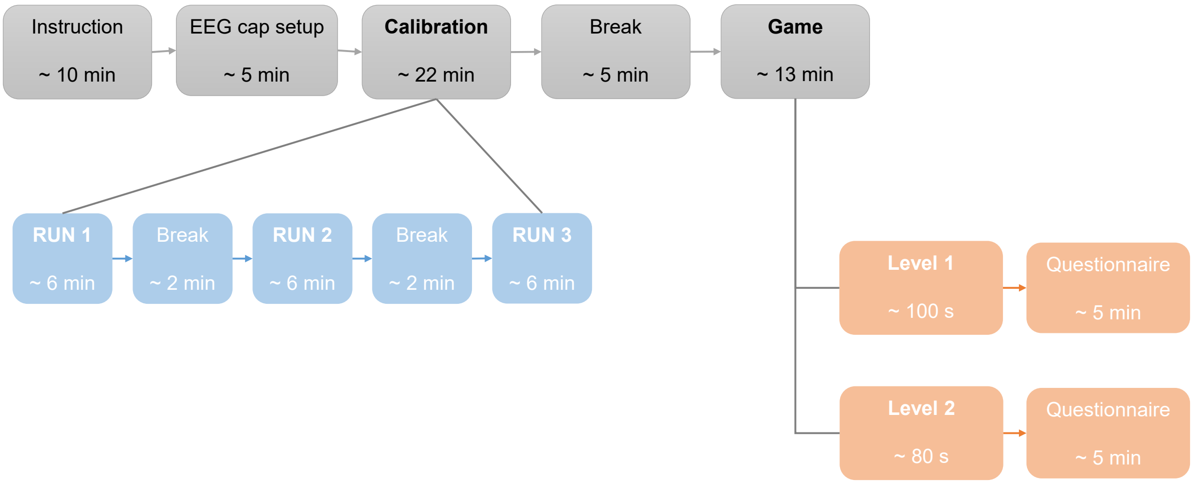


Figure 6.1: Schematic overview of the experimental session.

Although time consuming, a customized calibration process was necessary because EEG signals can vary significantly between subjects. Furthermore, variability within subjects was minimized by conducting both the calibration and the game phases on the same day in a controlled environment. This approach increases the reliability of the collected data and improves the training of the classification algorithm.

The calibration phase was conducted using a synchronous approach to identify the online classification model. Participants completed three runs, each requiring them to perform one of the following tasks:

- imagining the movement of their left hand;
- imagining the movement of their right hand;
- remain in an idle state.

Each run consisted of 45 trials, with 15 trials dedicated to each task. The duration of each run was approximately 6 min. The trial sequence was randomized to prevent participants from getting used to a specific order and anticipating the subsequent cue. In total, each participant completed 135 trials.

At the start of each trial, a fixation cross was displayed for 2.00 s to help the participants maintain a steady gaze. A visual cue, shown for 1.25 s along with the instruction "GET READY", indicated the task to be performed. This cue consisted of an arrow pointing to the right or left or a cross on the screen, indicating the motor imagery of the left or right hand or the state of relaxation, respectively. The participants then participated in the mental task for a period ranging from 3.00 s to 6.00 s without any visible physical movement. A brief randomized pause followed each trial to reduce automatic responses when starting mental imagery.

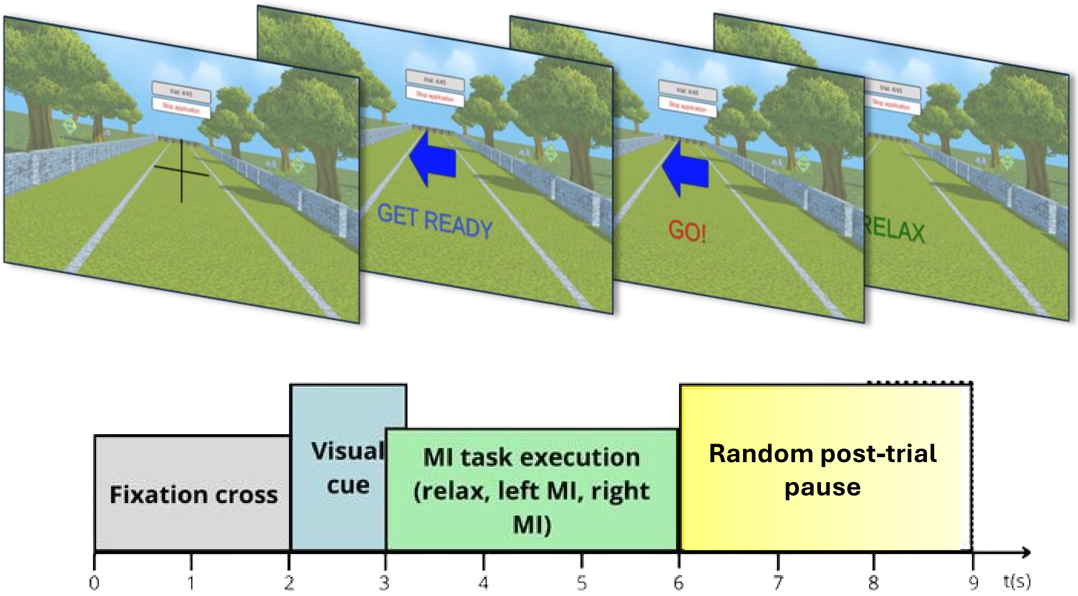


Figure 6.2: Example of a sequence of events for a single trial of the calibration.

The calibration phase was overseen using a custom application created in Unity that provided visual instructions to the participants (Fig. 6.2). The EEG data collected during this phase was processed offline. The aim was to train the processing algorithm that would be used in real-time during the subsequent game phase.

After calibration, participants engaged in two levels of a game that increased in difficulty. Throughout the game, real-time EEG signals controlled the horizontal movement of an avatar in a virtual environment. The main objective was to score as high as possible by collecting coins placed along a straight, three-lane path within a set time limit.

As the avatar continued to move forward, its horizontal position was dictated by the classification of EEG patterns, using the model developed during the calibration phase. Specifically, imagining a right-hand movement would shift the avatar to the right lane, imagining a left-hand movement would move it to the left lane, and no motor imagery would keep the avatar in its current lane.

The success of the game relied on how accurately the EEG signals during gameplay matched those recorded during the calibration. At the end of each game level, participants also filled out a self-assessment questionnaire to evaluate their engagement. Specifically, the questionnaire proposed in Chapter 5 was used. Using a 7-point Likert scale [215], the questionnaire inquired about:

- the perceived challenge of the task;

- the subjective perception of the time taken to complete the task;
- the emotional involvement evoked by the content of the task;
- the ease of maintaining concentration during the task;
- the visual attractiveness of the task stimuli.

6.2.2 Brain-computer interface architecture

The proposed architecture for the BCI system is illustrated in Fig. 6.3. It includes several essential components: a) a wearable, wireless, low-density EEG instrument for acquiring brain signals; b) a PC that runs MATLAB and Unity. The MATLAB software is used for processing the EEG signal, whereas Unity is used for creating the graphical interface and virtual game environment.

The EEG headset sends brain signal data to the PC through Bluetooth. A Simulink API is employed to capture the EEG data as it is generated to facilitate real-time signal acquisition. The communication between Simulink and Unity is established using the "User Datagram Protocol" (UDP) protocol. The setup allows for real-time signal processing in Simulink, which then sends the processed data to Unity. At the same time, Unity acts as a constant relay for real-time updates about the game's events and status, sending this information back to Simulink.

This architecture allows for a smooth integration of signal acquisition, processing, and interaction within the virtual environment, ensuring that the system functions in real time and provides an immersive, natural user experience.

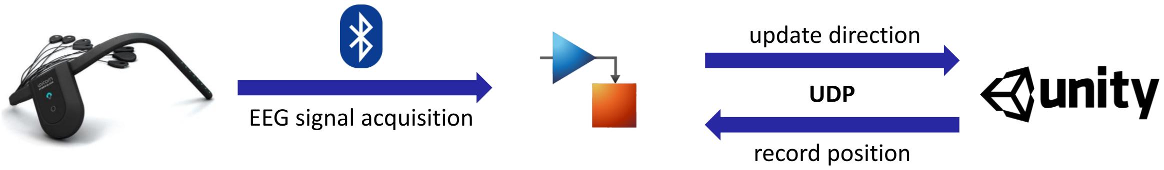


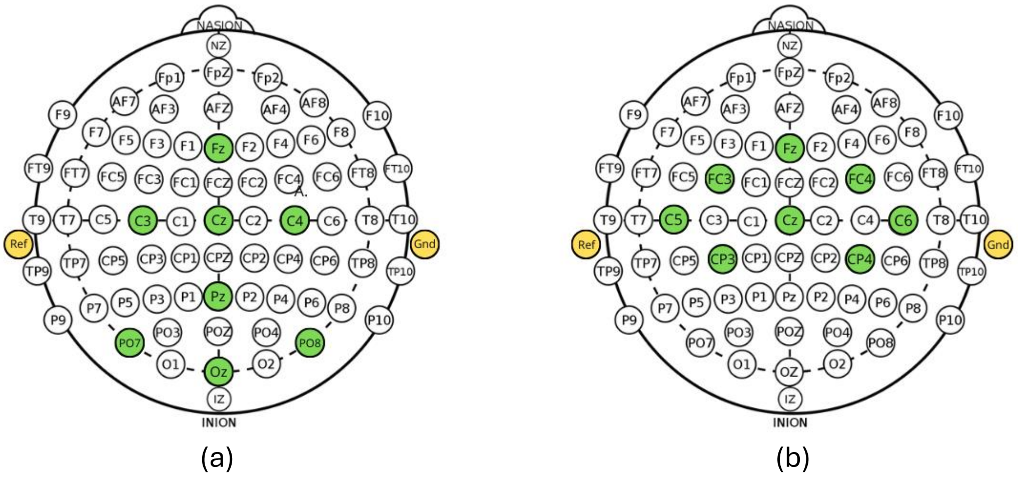
Figure 6.3: Block diagram of the proposed system.

Electroencephalographic instrument

The EEG instrument used in the experiments was the Unicorn Hybrid Black instrument (g.tec medical engineering GmbH, Austria), referred to as Device 1 in Chapter 2.

The Unicorn Hybrid Black features eight EEG acquisition channels. The electrodes are hybrid, allowing them to be used with or without conductive gel. For these experiments, conductive gel was applied to ensure high-quality signals and compensating for the instrument's limited spatial resolution.

The default configuration places the electrodes at Fz, C3, Cz, C4, Pz, PO7, Oz, and PO8. However, the configuration was modified for this proposal to enhance coverage of the motor imagery region of interest. The electrodes were positioned at Fz, FC3, FC4, C5, Cz, C6, CP3, and CP4. The difference between the two configuration is shown in



(a) (b)

Figure 6.4: Electrodes configuration for the EEG instrument. (a) Default configuration. (b) Proposed configuration.

Fig.6.4. Unlike other instrument, the Unicorn Hybrid Black allows for flexible electrode placement, enabling optimal customization for specific experimental needs.

The ground electrode was placed on the right mastoid, while the reference electrode was placed on the left mastoid using disposable surface adhesive electrodes. The EEG signal was acquired at a sampling frequency of 250 Hz with a resolution of 24 bits. Data was transmitted to a PC through the instrument's integrated Bluetooth interface.

To enable real-time processing, the instrument's Simulink API was used to directly receive EEG signals in Simulink. This eliminates the need to use the instrument's proprietary software and thus reduces the signal acquisition delay.

Processing pipeline

The signal processing pipeline used in this study was based on the "Filter Bank Common Spatial Pattern" (FBCSP) method [151], specifically tailored for a three-class classification problem. While deep learning techniques are often preferred in the literature for multi-class classification tasks due to their enhanced performance, this method was selected for its metrological reliability. Deep learning algorithms tend to show significant variability in their performance, which can make them less suitable for online processing scenarios.

The FBCSP approach was implemented in two separate phases: an offline phase and an online phase. Prior to each system use, the processing algorithm was calibrated using data from the current subject.

During the offline phase, EEG signals collected from a single trial were analysed using a sliding window of 2.00 s, which was shifted by 0.25 s. To ensure statistical reliability, a k-fold cross-validation method was employed to assess classification accuracy on a window-by-window basis. The data were randomly divided into k = 5 folds of equal size. The classification algorithm was executed k times, with one subset reserved for validation and the remaining subsets utilized for training.

EEG signals were filtered channel-by-channel into four standard non-overlapping frequency bands:

- 8 Hz to 12 Hz (μ-band);

- 12 Hz to 18 Hz (low β-band);
- 18 Hz to 28 Hz (high β-band);
- 28 Hz to 40 Hz Hz (low γ-band).

Chebyshev Type II filters with a band-stop attenuation of 20 dB were employed for this filtering process.

For each pair of classes (namely right hand versus left hand, right hand versus rest, and left hand versus rest), the "Common Spatial Patterns" (CSP) method was applied to create spatial filters. These filters improved the separability between the two classes by maximizing the variance of the EEG signals from one class while minimizing their variance in relation to the other classes. The processed EEG signals were then projected into a new vector space defined by the CSP filters using a linear transformation matrix. Three pairs of CSP filters were chosen for each frequency band, and the time-varying log-variances of each CSP-filtered epoch were calculated. To enhance the feature set, the "Minimal-Redundancy-Maximal-Relevance" (MRMR) method was used for feature selection [245]. The selected features were then classified using the "Linear Discriminant Analysis" (LDA). In particular, three binary classifiers were implemented:

- left versus rest (L vs. X);
- right versus rest (R vs. X);
- left versus right (L vs. R).

The goal of this analysis was to identify the best time window and feature set for training the algorithm for future online processing.

The online EEG data acquisition and processing module replicated the offline pipeline with some adjustments. It was developed in Simulink. EEG data epochs of 2.00 s were buffered and processed for each channel, with a temporal window shift of 0.25 s. Previous research has shown that a window length of 2.00 s yields optimal classification results [246]. Therefore, feedback on the avatar's movement was provided only after this interval had passed and been analysed.

The online processing pipeline, as shown in Fig.6.5, included the same stages as the offline pipeline. However, during the online phase, raw data were collected in real time while playing the game. An additional step in the online processing involved sending the outputs of the LDA classifiers to a moving average filter to smooth out short-term fluctuations and improve classifier reliability.

While the results from the offline classifiers could be evaluated independently, the online phase required a final decision to send commands to the avatar. The following decision rules were applied:

- if two classifiers indicated "right", the command was "move right";
- if two classifiers indicated "left", the command was "move left";
- in all other scenarios, the predicted class was "rest".

The avatar's movements were based on these predictions, providing real-time feedback to the user. This approach ensured robust and interpretable signal processing.

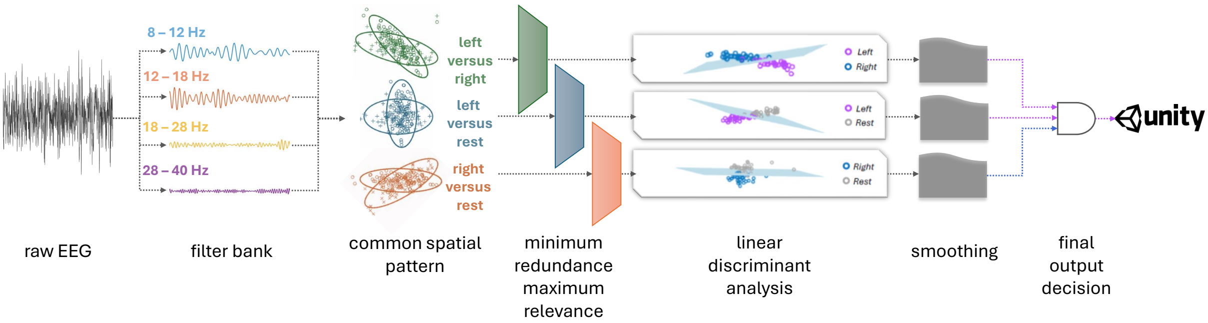


Figure 6.5: Online version of the proposed data processing pipeline.

6.2.3 Endless runner game

A Unity-based application was developed and used during the experiments. It used processed EEG signals to control an avatar and assess user performance during self-paced experiments.

The game application was created as a virtual space where a ghost-like avatar could move horizontally along a road divided into three lanes, flanked by walls. When launched, a connection message notified the user about the communication status with the Simulink model.


Waiting for Simulink connection...

Get ready!

You are about to begin in:

5,5

Completed

cted a total of 1 coin(s)

Figure 6.6: Sequence of scenes of the proposed game.

The game was designed with several scenes (Fig. 6.6). In the first scene, participants had the option to either start the game or exit. Choosing to exit would stop the game and end the Simulink execution. If the participants opted to play, a scene with the “GET READY!” message appeared for 10 s, along with a countdown to help participants prepare for the first level. The gameplay featured two levels that became progressively more challenging. Throughout the game, the score was shown in a red text box located in the top-right corner of the screen, while the elapsed time was displayed in the top-left corner. In the final scene, players could decide whether to replay the recently completed level, move on to the second level, or exit the game.

The endless feature of the game was achieved by repeatedly spawning a prefabricated game object. Each prefab contained the game floor along with additional features like trees, grass, walls, lanes, and coins.

For the first level, a total of 12 prefabs were pre-spawned. The first two of them were free of coins to allow the user to familiarize themselves with the game interface and focus on the motor imagery task required to collect the first coin. Coins were placed about 10 s apart, giving players enough time to navigate their avatar between each coin. This level had a fixed duration of 100 s.

In the second level, 14 prefabs were used. In this case, the first three prefabs were free of coins. The scrolling speed of the road increased dynamically based on the player's score, providing a more challenging experience. Coins were spaced 5 s apart, and the level duration was reduced to 80 s to reflect the increased effort required.

The user's goal was to achieve the highest score possible within a limited time by collecting coins scattered across the lanes. Each coin collected added 10 points to the score. Unlike most motor imagery-based BCI applications, the avatar moved continuously and autonomously, relying solely on motor imagery and relaxation states.

The avatar consisted in a 3D ghost prefab. It was modelled to facilitate collision handling with coins. Its cylindrical shape provided consistent and natural collision behavior, even when collecting coins laterally. An invisible game object, known as a tracker, was attached to the avatar to record game scores, level loading, coin collisions, and coin positions.

The game was built on the basis of continuous two-way communication between Unity and Simulink to integrate user-generated EEG data with the game mechanics. Simulink processed EEG signals in real time, classifying motor imagery tasks and sending the resulting labels to Unity. These labels controlled the avatar's movements. In turn, Unity sent game-related information back to Simulink, which included details like the current level, the avatar's position, the timing and locations of coin appearances (in lanes 1, 2, or 3), and whether coins were collected or missed. This back-and-forth communication ensured real-time synchronization between the game environment and the BCI system. At the end of each game session, Simulink logged all received data, including EEG signals, classified labels, and Unity's gameplay data.

However, traditional game metrics such as the in-game score and the number of collected coins were not enough to qualitatively assess participants' ability to control the BCI system. The random placement of coins across different lanes added variability, meaning that a participant could gather many coins even if they were limited to one lane due to a control bias.Conversely, participants who demonstrated better control over the avatar might narrowly miss coins due to their spatial placement, resulting in lower scores that did not accurately reflect their performance.

To overcome these challenges, a new metric called "coin error" was introduced to objectively measure how well participants controlled the avatar. This metric used real-time positional data of the avatar and the locations where coins appeared to determine the minimum Euclidean distance between the avatar and each coin during the game. If a coin was collected, the corresponding distance was set to zero. Unlike traditional trajectory-based error metrics, such as the "root mean square error" (RMSE), which assess deviation from an ideal path without considering task-specific objectives, the "coin error" metric focuses on performance at key decision points rather than evaluating the entire trajectory. By integrating task-related goals, it provides a more informative assessment of control.

A lower "coin error" value indicated a closer approach to the coins, indicating better voluntary control over the system. This metric enabled a more comprehensive evaluation by taking into account not just the coins collected but also how closely missed coins were approached. The metric was mathematically expressed as:

$$coin\ error = \frac{1}{max\ error}\frac{1}{N}\sum_{j=1}^{N}\sqrt{(x_{coin,j} - x_{avatar,j})^2 + (y_{coin,j} - y_{avatar,j})^2} \tag{6.1}$$

where *max error* represents the maximum possible mean distance between the ideal coin path and the avatar's path, N is the total number of coins spawned during the level, x_{coin_n}, y_{coin_n} denote the position of the j-th coin, and x_{avatar}, y_{avatar} represent the avatar's position at the closest point to the j-th coin.

6.2.4 Post-hoc analysis

The relationship between calibration phase performance and the "coin error" metric was investigated by analysing both the three binary classification accuracies (left hand vs. right hand, left hand vs. rest, and right hand vs. rest) and the composite accuracy. The composite accuracy was intended to estimate real-time classification accuracy using offline data, mimicking the decision-making process used in online gameplay. First, the optimal 2-second time window and the corresponding number of features were determined. Then, a 10-fold cross-validation was used to derive the composite accuracy from binary classifications. To assess correlation, Pearson's correlation test was used for normal distributed and homoscedastic data. For data that did not follow a normal distribution, Spearman's correlation test was used. The Shapiro-Wilk test [223] was used to test the normality of the data. The F-test [247] was used to assess homoscedasticity, ensuring equal variance across groups.

Finally, the responses to the questionnaire were analysed. The median for each question was calculated among all participants. Statistical tests were conducted to compare responses between the first and second level questionnaires. Since the same participants responded under two different conditions, the two distributions were not independent. After normality was assessed, a paired t-test or a Wilcoxon signed rank test was performed to identify significant differences.

6.2.5 Results and discussions

Participants

Twenty-three healthy participants (S01–S23) took part in the experiments. The group consisted of seven females and sixteen males, all with normal or corrected-to-normal vision. Their ages ranged from 18 to 32 years (25 ± 7 years, mean and standard deviation). Except for one participant (S06), all were unfamiliar with BCI technology.

The study was conducted at the Augmented Reality for Health Monitoring Laboratory (ARHeMLab) of the University of Naples Federico II, Italy. Ethical approval was obtained from the Ethics Committee for Psychological Research of the Department of Humanities, University of Naples Federico II (protocol no. 24/2024, approved on 15/07/2024). Informed consent was collected from all participants before the experiments.

Offline performance

The EEG data collected during the calibration phase was used to identify the best time window and the number of features needed for training the online processing algorithm. Collecting data individually for each subject was necessary due to inter-subject variability in brain signals.

Fig. 6.7 shows how cross-validation accuracy changes over time for a representative subject (S06) across the three task pairs. A sliding window of 2.00 s with a 0.25-second shift was used to assess how classification accuracy changed over time. The blue bounded

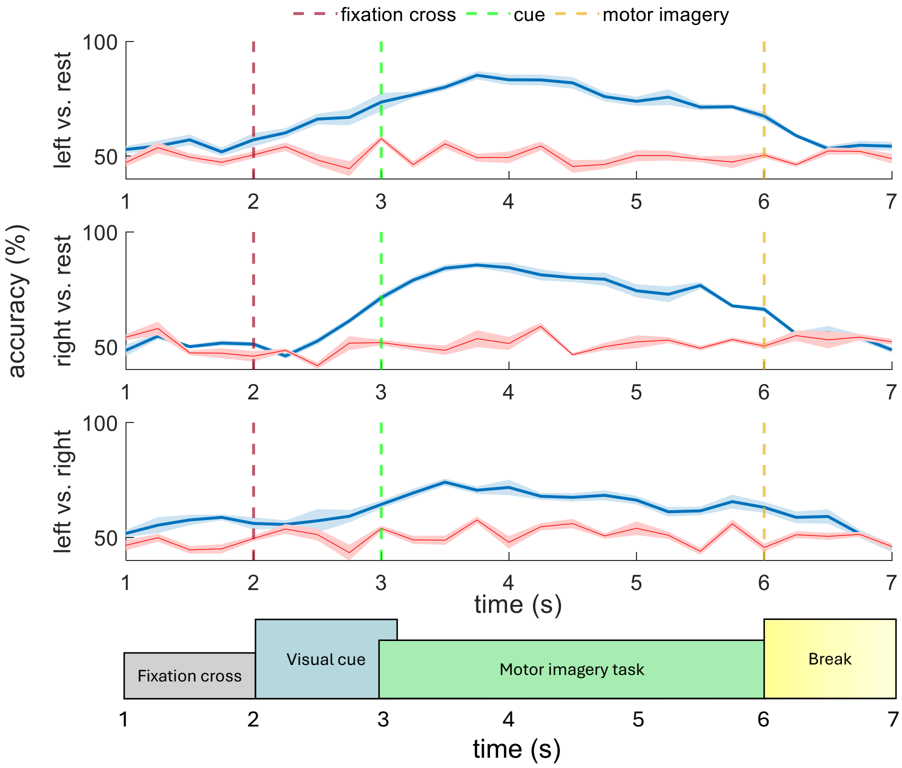


Figure 6.7: Cross-validation accuracy over time for a representative subject (S06) across the three classification tasks. The blue bounded line represents the mean classification accuracy with its standard deviation, while the red bounded line corresponds to the permutation test results. Vertical markers indicate key time instants (cue onset, task start, and task end).

line indicates the average classification accuracy along with its standard deviation, while the red bounded line reflects the results of the permutation test, which estimates chance-level performance. The permutation test curve hovers around 50 %, suggesting a random classification outcome, while the actual accuracy curve diverges from this baseline. This is especially true during the time window related to the execution of the motor imagery task. Key time points are also highlighted: cue onset (yellow), task initiation (green), and task termination with the onset of the relaxation phase (orange). The highest classification accuracy was mainly observed within the time window corresponding to the execution of the task. For each subject, the 2.00-second window that produced the best accuracy was chosen for training the online classifier. Overall, the highest classification accuracies were achieved with up to 25 features, beyond which accuracy values remained approximately stable across all classifiers.

Tab. 6.1 shows the cross-validation classification accuracies along with their standard deviations for each subject. They were assessed using the optimal time window and number of features for each subject. The results are presented for binary classifications (left vs. rest, right vs. rest, left vs. right) as well as for the composite classification. The highest classification accuracies were recorded for subject S06 (88 ± 13 % for left vs. rest and 88 ± 11 % for right vs. rest) and subject S19 (84 ± 10 % for left vs. right). S06 also achieved the highest composite accuracy (73 ± 14 %). On average, the classification accuracies for lef vs. rest and right vs. rest were greater than that for left vs. right.

Instead, the composite accuracy was 54 ± 15 %, exceeding the theoretical chance level of 33 %. Notably, the composite accuracy reflects the decision-making process used during online classification, where all three classes are evaluated at the same time.

Table 6.1: Cross-validation accuracies calculated for each participant using calibration data. Both binary and composite accuracy are reported.

Participants	**Offline phase accuracies (%)**			
	left vs. rest	**right vs. rest**	**left vs. right**	**composite**
S01	69 ± 11	74 ± 10	69 ± 13	55 ± 16
S02	72 ± 12	78 ± 10	68 ± 11	53 ± 18
S03	63 ± 11	61 ± 9	61 ± 8	43 ± 15
S04	68 ± 11	69 ± 14	64 ± 13	53 ± 15
S05	71 ± 11	65 ± 12	65 ± 12	44 ± 14
S06	88 ± 13	88 ± 11	77 ± 12	73 ± 14
S07	68 ± 12	68 ± 9	61 ± 11	46 ± 14
S08	85 ± 8	84 ± 9	70 ± 10	65 ± 16
S09	87 ± 10	87 ± 11	76 ± 9	66 ± 15
S10	67 ± 9	73 ± 12	70 ± 10	53 ± 16
S11	86 ± 11	84 ± 11	64 ± 11	58 ± 15
S12	73 ± 11	71 ± 10	63 ± 11	55 ± 17
S13	71 ± 10	67 ± 11	57 ± 12	47 ± 15
S14	86 ± 12	89 ± 12	69 ± 10	68 ± 16
S15	78 ± 9	67 ± 13	64 ± 11	50± 15
S16	61 ± 11	64 ± 12	60 ± 11	39 ± 16
S17	65 ± 10	68 ± 11	63 ± 11	44 ± 15
S18	77 ± 12	78 ± 10	63 ± 11	54 ± 15
S19	70 ± 9	74 ± 11	84 ± 10	54 ± 17
S20	77 ± 8	77 ± 10	72 ± 12	59 ± 16
S21	81 ± 12	75 ± 10	66 ± 10	57 ± 15
S22	62 ± 11	70 ± 9	64 ± 11	50 ± 15
S23	63 ± 9	68 ± 12	72 ± 10	53 ± 16
average	73 ± 11	73 ± 11	67 ± 11	54 ± 15

Fig. 6.8 illustrates these results graphically, making it easier to identify relationships in classification accuracies. In particular, Fig. 6.7(a) shows the classification accuracies, while Fig. 6.7(b)b highlights the calibration performance using Cohen's kappa coefficient [162]. This kappa coefficient takes into account chance-level performance, allowing for a more straightforward comparison between binary classification results and the three-class composite accuracy. Interestingly, the left vs. right classification accuracy aligns closely with the composite accuracy.

These findings indicate a greater separability between motor imagery tasks and the resting state, which is consistent with previous research [45, 248, 229]. However, the composite classification accuracy remains constrained, similarly to left vs. right classification performance. This limitation may arise from: i) the classification algorithm's ability to effectively distinguish features; ii) the limited experience of the participants with motor imagery tasks, which could hinder their ability to produce distinct neural activity patterns.

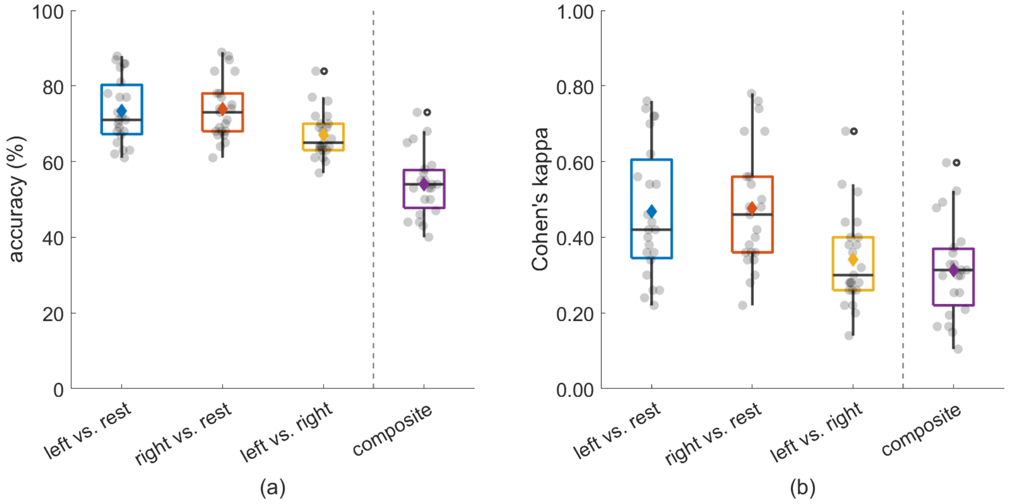


Figure 6.8: Cross-validation results across subjects. (a) Classification accuracies for binary and composite tasks. (b) Cohen's kappa coefficients accounting for chance-level performance. Diamond markers indicate mean values across subjects.

Some alternative approaches were explored offline, including a well-established reference pipeline for the FBCSP method [151] and deep learning techniques. The pipeline described in [151] is considered a benchmark for analysing motor imagery EEG data and differs from the approach taken in this study in two main ways:

- EEG pre-processing employs nine non-overlapping Chebyshev type II filters within the range from 4 Hz to 40 Hz;
- "Mutual Information-based Best Individual Feature" algorithm is used as feature extraction method to identify CSP features that exhibit the highest mutual information with the class labels.

In addition, two deep learning models were tested:

- EEGNet [249], a compact convolutional neural network that usess depthwise and separable convolutions;
- SincEEGNet [250], an EEGNet variant that integrates Sinc filters for direct extraction of frequency-specific features, thereby reducing the number of learnable parameters.

For both models, the training parameters were established based on the reference works [249, 250], using 32 filters in the initial convolutional block and a fixed seed of 45 to guarantee reproducibility.

The goal of the comparative analysis was to find the most effective method for achieving high classification accuracy while also being computationally feasible for real-time use. For deep learning models, classification was performed directly on the three-class problem (left, right, rest). This is in line with the evidence that neural networks tend to achieve superior performance in multiclass classification tasks [16]. To reduce the risk of overfitting, a "Leave-One-Subject-Out" (LOSO) approach was employed for training. In this method, the data from each subject were reserved solely for testing, while the model was trained on the data from all other subjects. This setup mimics a zero-calibration scenario, which is consistent with applications where the system should generalize to new users without subject-specific training.

The results of the comparative evaluation are shown in Fig. 6.9. Classification accuracies were represented using boxplots, with diamond markers indicating mean values. For approaches that do not use deep learning, both binary classification accuracies and overall composite classification accuracy were reported. In the case of deep neural networks, only the multiclass classification accuracy was reported.

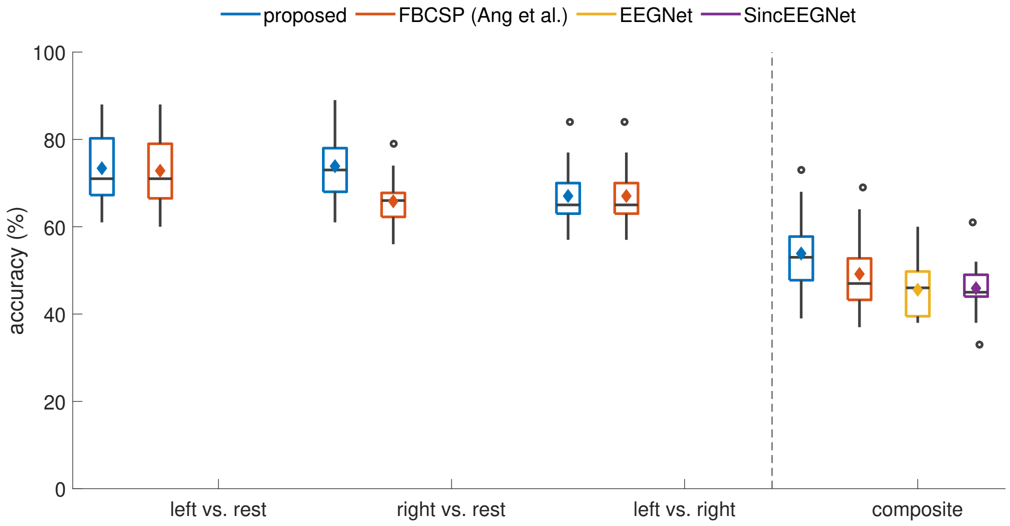


Figure 6.9: Comparison of classification accuracies across different approaches. Diamond markers indicate mean values. Non-deep learning approaches include binary and composite accuracies, while deep learning methods report only composite accuracies.

The results show that the pipeline used in this study is more effective than the FBCSP method for the right vs. rest classification task and in composite classification. Moreover, the proposed approach has less computational cost than [151] due to the less number of bandpass filters employed. Additionally, it performs better than deep neural networks, which struggle to generalize in this context. Although EEGNet and SincEEGNet have demonstrated high performance on benchmark datasets [249, 250], their applicability in data-limited scenarios and under zero-calibration conditions is reduced.

Online performance

Tab. 6.2 shows how the subjects performed during the gaming phase, detailing both the proposed "coin error" metric and the percentage of coins collected by each subject in Level 1 and Level 2. Some subjects did not take part in Level 2, so their entries in the table are left blank.

A relationship can be observed between the number of collected coins and the "coin error" metric can be observed. However, "coin error" provides a more comprehensive measure of avatar control, as previously introduced in [251]. In some cases, the participants had low "coin error" values. This means that they effectively controlled the avatar, even though they collected only a few coins. For example, subject S07 in Level 1 and A16 in Level 2 collected just 33 % and 19 % of the coins, respectively, yet demonstrated

Table 6.2: Online performance. Both coin error and percentage of collected coins are reported for each level.

Participants	**Level 1**		**Level 2**	
	coin error	**collected coins (%)**	**coin Error**	**collected coins (%)**
S01	0.25	44	0.37	50
S02	0.31	44	0.42	53
S03	0.47	11	0.38	13
S04	0.56	11		
S05	0.40	22	0.45	35
S06	0.06	66	0.23	37
S07	0.29	33	0.44	29
S08	0.26	44	0.48	13
S09	0.29	33	0.23	38
S10	0.55	0	0.57	25
S11	0.35	33		
S12	0.42	11	0.45	31
S13	0.33	33	0.36	47
S14	0.41	44	0.32	47
S15	0.51	33	0.41	35
S16	0.65	11	0.31	19
S17	0.62	22	0.52	35
S18	0.27	33		
S19	0.32	33		
S20	0.50	22	0.56	24
S21	0.40	22		
S22	0.20	33	0.24	38
S23	0.55	33		

good control with “coin error” values of 0.29 and 0.31. On the other hand, some subjects gathered a larger number of coins—possibly by chance—while showing poor control. For instance, subject S23 in Level 1 and S17 in Level 2 collected 33 % and 35 % of the coins, but had higher “coin error” values of 0.55 and 0.52, respectively. Fig. 6.10 provides an example of the gaming performance of subject S06 on both levels. In this figure:

- the red path shows the actual trajectory taken by the avatar;
- the black path illustrates the ideal trajectory, linking consecutive coin positions;
- yellow coins represent the coins collected, while gray coins indicate those that were missed.

This visualization emphasizes how the “coin error” metric captures deviations from the optimal path, offering a quantitative measure of avatar control. Notably, there are times when the avatar comes close to a coin but fails to collect it, highlighting that mere proximity does not guarantee successful coin collection. This underscores the importance of the “coin error” metric in assessing trajectory accuracy beyond simple success rates.

To further validate the reliability of the control signals used during the gaming phase, an additional analysis to evaluate the potential impact of ocular and muscular artifacts on

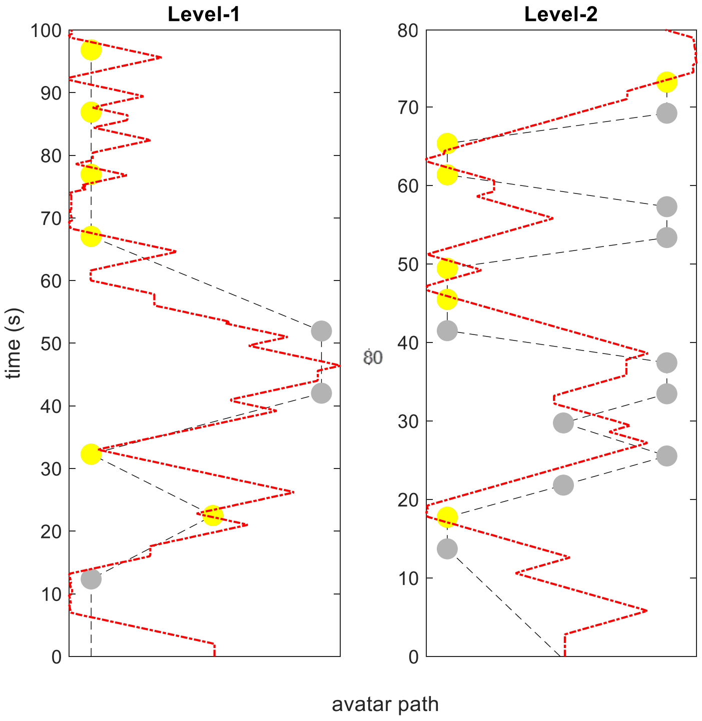


Figure 6.10: Example of avatar trajectory for subject S06 in Level 1 and Level 2. Red path: actual trajectory followed by the avatar. Black path: ideal trajectory connecting consecutive coins. Yellow coins: successfully collected coins. Gray coins: missed coins.

gameplay was conducted. As found in Chapter 3, the “Artifact Subspace Reconstruction” (ASR) method effectively corrects artifacts up to five EEG channels. Therefore, ASR was used to verify the integrity of the neural signals used for control.

Fig. 6.11 shows the EEG signals acquired during Level 1 for subject S06. The raw signal filtered within the 8 Hz to 40 Hz band was compared with the same signal after ASR-based artifact removal. The figure consists of two subplots: the left figure shows the complete 100-second recording, while the right figure offers a zoomed-in view of a 2-second segment. The near-complete overlap between the two signals throughout the level shows that artifact contamination did not affect the neural features used for control. The “relative root mean square error” (rRMSE) between the two signals was also calcutated. It resulted on the order of 10^{-6}, indicating minimal deviation.

Post-hoc analysis results

The relationship between online and offline performance was investigated. Fig. 6.12 shows the relationship between the classification accuracies obtained offline and the “coin error” metric in Level 1. The goal was to identify any potential correlations. Each subplot corresponds to a different classification pair (i.e., left vs. rest, right vs. rest, left vs. right) along with the composite accuracy. Given the non-homoscedastic nature of the data, Spearman’s rank correlation coefficient was chosen over Pearson’s correlation test. For each case, the Spearman’s rank correlation coefficient (ρ) and the corresponding p-value are reported.

The results indicated a significant correlation for the right-rest discrimination task

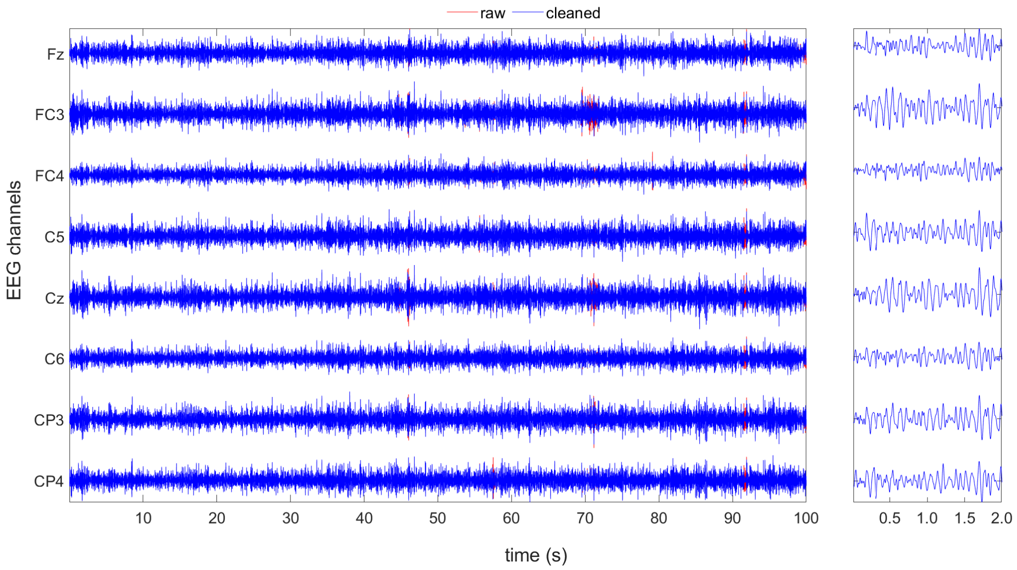


Figure 6.11: EEG signals recorded during Level 1 for subject S06. The raw signal (red) and the ASR-cleaned signal (blue) are shown. The left figure displays the full 100-second recording, while the right figure provides a 2-second zoom.

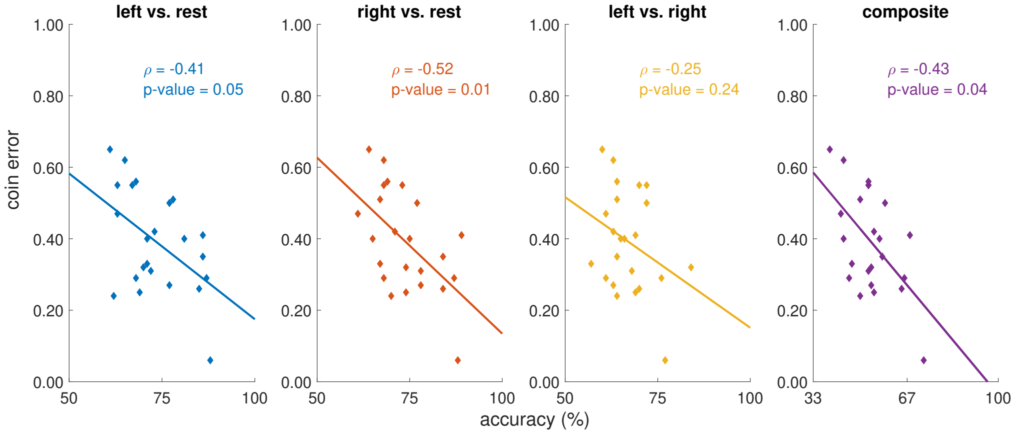


Figure 6.12: Scatter plots show the relationship between the classification accuracies obtained offline and the "coin error" metric in Level 1 for each classification pair along with the composite accuracy. For each case, the Spearman's rank correlation coefficient (ρ) and the corresponding p-value are reported.

and the composite classification of three classes. However, no significant correlation was found for the left-rest and left-right discrimination tasks. This suggests that participants with more distinct neural patterns during calibration tend to achieve better avatar control in the gaming phase. This trend is further supported by the presence of a clear outlier in all plots of Fig. 6.12, corresponding to subject S06, who achieved the lowest CE value. It is important to note that S06 is a trained participant with proven skills in motor imagery tasks. The lack of correlation for left-right discrimination highlights the inherent challenges in differentiating between left- and right-hand motor imagery, which affects the reliability of classification and ultimately influences gaming performance.

Since the performance of motor imagery-based BCIs is closely tied to how engaged users are in the task, an additional analysis was conducted using self-reported engagement scores. Fig. 6.13 shows the results from a post-experiment questionnaire, where median scores for each engagement dimension were calculated. The focus was on participants who completed both levels.

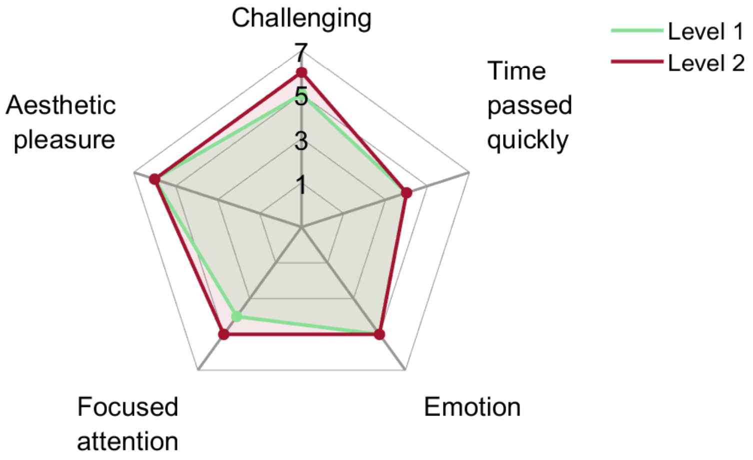


Figure 6.13: Comparison of self-reported engagement scores across different dimensions for Level 1 and Level 2. Median values were computed for participants who completed both levels.

In general, a slight differences in engagement scores between the two levels was found. However, Level 2 showed higher scores in the “Challenging” and “Focused attention” dimensions. The Wilcoxon signed-rank test indicated that these differences were not statistically significant. This suggests that participants did not find the increased difficulty in Level 2 to be overly frustrating or disengaging. Rather, the adaptive difficulty mechanism in Level 2 effectively maintained a good balance between challenge and engagement, ensuring consistent cognitive involvement throughout the task.

Conclusions

This doctoral thesis investigated the metrological challenges of wearable brain-computer interfaces, focusing on their transition from research laboratories to real-world applications. Despite their growing popularity for diverse applications, the performance of these interfaces remains a substantial constraint. This problem is particularly evident in the context of non-invasive, consumer-grade EEG systems.

The research focused specifically on motor imagery-based brain-computer interfaces, where users control the system through imagined movements rather than actual physical actions. The aim was to improve them from a number of perspectives, namely working on acquisition delay, artefact removal, processing, and user engagement detection. The latter is particularly important for motor imagery-based systems, as they rely entirely on voluntary brain activity. In contrast to stimulus-driven interfaces, their performance is strongly influenced by the user's concentration.

Chapter 1 introduced the basic principles of brain-computer interface technology, categorizing systems by their level of invasiveness and operational methods. It highlighted electroencephalography as a non-invasive way to measure brain activity. Important metrological factors, such as spatial resolution and signal-to-noise ratio, were discussed, especially concerning low-density electroencephalographic instrumentations that enhance wearability. The chapter underscored the importance of the motor imagery paradigm and the metrological challenges associated with its measurement and evaluation.

Chapter 2 examined the acquisition delay in wireless electroencephalographic instrumentations. A dedicated measurement station was proposed to assess this delay. The impact of delay on the detection of time-locked electroencephalographic phenomena was explored. The results indicated that both the mean acquisition delay and its associated uncertainty adversely affect electroencephalographic-based measurements. While the mean delay can be corrected after characterization, the variability in acquisition presents a significant source of uncertainty. This is especially challenging for detecting low signal-to-noise ratio events. These findings highlight the need to minimize both the magnitude and uncertainty of the delay to improve the reliability of the measurement and the performance of the system.

Chapter 3 focused on artifact contamination in electroencephalographic signals. This is a critical issue, especially in low-density setups. A comparative analysis of artifact removal techniques was conducted to identify the most effective method for low-channel configurations. Then, a new hybrid artifact removal method specifically designed for low-density setups was proposed. Unlike existing methods, it aims to eliminate multiple types of artifacts while processing all available channels together to retain as much information as possible. The study emphasized the importance of validating artifact removal techniques using real data, despite the absence of a pure electroencephalographic ground truth. Synthetic datasets can support this process, but they should not replace evaluations on real acquisitions.

Chapter 4 presented a metrological analysis of advanced electroencephalographic processing techniques for motor imagery. The evaluation emphasized key factors such as accuracy, uncertainty, repeatability, and reproducibility. It identified limitations in existing methodologies. In particular, the variability in motor tasks, electroencephalographic channel configurations, and participant samples. To enhance the reproducibility of results, it is crucial to standardize experimental procedures and performance assessment metrics. The findings underscore the necessity for consistent validation protocols to ensure measurement consistency across different studies.

Chapter 5 investigated the estimation of engagement based on real-time electroencephalographic data. A classification model was developed to differentiate brain activity patterns associated with engaging tasks from those during resting states. After training, the model was able to provide continuous engagement estimates across various applications. The results showed that the engagement estimates were in line with the levels of self-reported engagement of the users, confirming the effectiveness of the model.

Chapter 6 combined insights from the previous chapters into a real-time application. An endless runner game controlled by a low-density electroencephalography-based brain-computer interface was developed. Offline analyses demonstrated the ability to distinguish motor imagery tasks from resting states, with higher classification accuracies for left hand motor imagery vs. rest and right hand motor imagery vs. rest compared to left vs. right hand motor imageries. Real-time performance was evaluated based on the number of coins collected and a custom “coin error” metric. This metric provided deeper insight into performance by considering the proximity to coins rather than solely counting collected ones. Calibration accuracies for right hand motor imagery vs. rest and composite classifications were found to correlate with the “coin error” metric, highlighting the relationship between offline and real-time motor imagery abilities. A user questionnaire indicated that the adaptive game design effectively maintained engagement while reducing frustration, emphasizing the significance of difficulty progression in brain-computer interfaces applications.

In synthesis, this thesis provided key insights into the metrological challenges and technological advancements required for the implementation of a practical brain-computer interface. There are trade-offs between reducing acquisition delays and incorporating real-time artifact removal, which adds extra processing latency. Similarly, a balance must be found between system wearability and the range of detectable neural phenomena.

Bibliography

[1] Aleksandra Kawala-Sterniuk et al. “Summary of over fifty years with brain-computer interfaces—a review”. In: *Brain Sciences* 11.1 (2021), p. 43.

[2] Jonathan R Wolpaw et al. “Brain–computer interfaces for communication and control”. In: *Clinical neurophysiology* 113.6 (2002), pp. 767–791.

[3] Remigiusz J Rak, Marcin Kołodziej, and Andrzej Majkowski. “Brain-computer interface as measurement and control system the review paper”. In: *Metrology and Measurement Systems* (2012), pp. 427–444.

[4] BCI SOCIETY. `https://bcisociety.org/bci-definition/`.

[5] Jonathan R Wolpaw and E Winter Wolpaw. “Brain-computer interfaces: something new under the sun”. In: *Brain-computer interfaces: principles and practice* 14 (2012).

[6] Rabie A Ramadan and Athanasios V Vasilakos. “Brain computer interface: control signals review”. In: *Neurocomputing* 223 (2017), pp. 26–44.

[7] Simanto Saha et al. “Progress in brain computer interface: Challenges and opportunities”. In: *Frontiers in systems neuroscience* 15 (2021), p. 578875.

[8] Zhi-Ping Zhao et al. “Modulating brain activity with invasive brain–computer interface: A narrative review”. In: *Brain Sciences* 13.1 (2023), p. 134.

[9] Alexandre Moly et al. “An adaptive closed-loop ECoG decoder for long-term and stable bimanual control of an exoskeleton by a tetraplegic”. In: *Journal of Neural Engineering* 19.2 (2022), p. 026021.

[10] Mamunur Rashid et al. “Current status, challenges, and possible solutions of EEG-based brain-computer interface: a comprehensive review”. In: *Frontiers in neurorobotics* 14 (2020), p. 515104.

[11] Jürgen Mellinger et al. “An MEG-based brain–computer interface (BCI)”. In: *Neuroimage* 36.3 (2007), pp. 581–593.

[12] Noman Naseer and Keum-Shik Hong. “fNIRS-based brain-computer interfaces: a review”. In: *Frontiers in human neuroscience* 9 (2015), p. 3.

[13] Bing Du et al. “fmri brain decoding and its applications in brain–computer interface: A survey”. In: *Brain Sciences* 12.2 (2022), p. 228.

[14] Ganesh Pandarinathan et al. “The potential of cognitive neuroimaging: A way forward to the mind-machine interface”. In: *Journal of Imaging* 4.5 (2018), p. 70.

[15] Pasquale Arpaia et al. *Wearable Brain-Computer Interfaces: Prototyping EEG-Based Instruments for Monitoring and Control.* CRC Press, 2023.

[16] Pasquale Arpaia et al. “How to successfully classify EEG in motor imagery BCI: a metrological analysis of the state of the art”. In: *Journal of Neural Engineering* 19.3 (2022), p. 031002.

[17] Aurora Saibene et al. “EEG-based BCIs on motor imagery paradigm using wearable technologies: a systematic review”. In: *Sensors* 23.5 (2023), p. 2798.

[18] Upasana Talukdar, Shyamanta M Hazarika, and John Q Gan. “Motor imagery and mental fatigue: inter-relationship and EEG based estimation”. In: *Journal of computational neuroscience* 46 (2019), pp. 55–76.

[19] Khalida Douibi et al. “Toward EEG-based BCI applications for industry 4.0: Challenges and possible applications”. In: *Frontiers in Human Neuroscience* 15 (2021), p. 705064.

[20] Teng Cao et al. “Objective evaluation of fatigue by EEG spectral analysis in steady-state visual evoked potential-based brain-computer interfaces”. In: *Biomedical engineering online* 13 (2014), pp. 1–13.

[21] Aldo Mora-Sánchez et al. “A brain–computer interface for the continuous, real-time monitoring of working memory load in real-world environments”. In: *Cognitive Neurodynamics* 14 (2020), pp. 301–321.

[22] Boris Reuderink, Anton Nijholt, and Mannes Poel. “Affective Pacman: A frustrating game for brain-computer interface experiments”. In: *Intelligent Technologies for Interactive Entertainment: Third International Conference, INTETAIN 2009, Amsterdam, The Netherlands, June 22-24, 2009. Proceedings 3*. Springer. 2009, pp. 221–227.

[23] Angela Natalizio et al. “Real-time estimation of EEG-based engagement in different tasks”. In: *Journal of Neural Engineering* 21.1 (2024), p. 016014.

[24] Maryam Alimardani and Kazuo Hiraki. “Passive brain-computer interfaces for enhanced human-robot interaction”. In: *Frontiers in Robotics and AI* 7 (2020), p. 125.

[25] Hans Berger. “Über das Elektrenkephalogramm des Menschen. XIV.” In: *Archiv für Psychiatrie und Nervenkrankheiten* (1938).

[26] Jacques J Vidal. “Toward direct brain-computer communication”. In: *Annual review of Biophysics and Bioengineering* 2.1 (1973), pp. 157–180.

[27] Valer Jurcak, Daisuke Tsuzuki, and Ippeita Dan. “10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems”. In: *Neuroimage* 34.4 (2007), pp. 1600–1611.

[28] Guangli Li, Sizhe Wang, and Yanwen Y Duan. “Towards gel-free electrodes: A systematic study of electrode-skin impedance”. In: *Sensors and Actuators B: Chemical* 241 (2017), pp. 1244–1255.

[29] Guang-Li Li et al. “Review of semi-dry electrodes for EEG recording”. In: *Journal of Neural Engineering* 17.5 (2020), p. 051004.

[30] Pasquale Arpaia et al. “Paving the Way for Motor Imagery-Based Tele-Rehabilitation through a Fully Wearable BCI System”. In: *Sensors* 23.13 (2023), p. 5836.

[31] Jiawei Xu et al. “Active electrodes for wearable EEG acquisition: Review and electronics design methodology”. In: *IEEE reviews in biomedical engineering* 10 (2017), pp. 187–198.

[32] Michal Teplan et al. “Fundamentals of EEG measurement”. In: *Measurement science review* 2.2 (2002), pp. 1–11.

[33] Aruna Tyagi, Sunil Semwal, and Gautam Shah. “A review of EEG sensors used for data acquisition”. In: *Journal of Computer Applications (IJCA)* (2012), pp. 13–17.

[34] W David Hairston et al. “Usability of four commercially-oriented EEG systems”. In: *Journal of neural engineering* 11.4 (2014), p. 046018.

[35] Julia WY Kam et al. “Systematic comparison between a wireless EEG system with dry electrodes and a wired EEG system with wet electrodes”. In: *NeuroImage* 184 (2019), pp. 119–129.

[36] Kaido Värbu, Naveed Muhammad, and Yar Muhammad. “Past, present, and future of EEG-based BCI applications”. In: *Sensors* 22.9 (2022), p. 3331.

[37] Yu Xie and Stefan Oniga. “A review of processing methods and classification algorithm for EEG signal”. In: *Carpathian Journal of Electronic and Computer Engineering* 13.1 (2020), pp. 23–29.

[38] Drishti Yadav, Shilpee Yadav, and Karan Veer. “A comprehensive assessment of Brain Computer Interfaces: Recent trends and challenges”. In: *Journal of Neuroscience Methods* 346 (2020), p. 108918.

[39] Gert Pfurtscheller and Christa Neuper. “Motor imagery and direct brain-computer communication”. In: *Proceedings of the IEEE* 89.7 (2001), pp. 1123–1134.

[40] Ujwal Chaudhary, Niels Birbaumer, and Ander Ramos-Murguialday. “Brain–computer interfaces for communication and rehabilitation”. In: *Nature Reviews Neurology* 12.9 (2016), pp. 513–525.

[41] Muhammad Ahmed Khan et al. “Review on motor imagery based BCI systems for upper limb post-stroke neurorehabilitation: From designing to application”. In: *Computers in biology and medicine* 123 (2020), p. 103843.

[42] Giovanna Cuomo et al. “Motor imagery and gait control in Parkinson’s disease: Techniques and new perspectives in neurorehabilitation”. In: *Expert Review of Neurotherapeutics* 22.1 (2022), pp. 43–51.

[43] Elisabetta Sarasso et al. “Action observation and motor imagery improve dual task in Parkinson’s disease: a clinical/fMRI study”. In: *Movement Disorders* 36.11 (2021), pp. 2569–2582.

[44] Dennis J McFarland et al. “Mu and beta rhythm topographies during motor imagery and actual movements”. In: *Brain topography* 12 (2000), pp. 177–186.

[45] Gert Pfurtscheller and FH Lopes Da Silva. “Event-related EEG/MEG synchronization and desynchronization: basic principles”. In: *Clinical neurophysiology* 110.11 (1999), pp. 1842–1857.

[46] Evan M Gordon et al. “A somato-cognitive action network alternates with effector regions in motor cortex”. In: *Nature* 617.7960 (2023), pp. 351–359.

[47] Wilder Penfield and Edwin Boldrey. “Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation”. In: *Brain* 60.4 (1937), pp. 389–443.

[48] Amardeep Singh et al. “A comprehensive review on critical issues and possible solutions of motor imagery based electroencephalography brain-computer interface”. In: *Sensors* 21.6 (2021), p. 2173.

[49] Pasquale Arpaia et al. “Metrological characterization of consumer-grade equipment for wearable brain–computer interfaces and extended reality”. In: *IEEE Transactions on Instrumentation and Measurement* 71 (2021), pp. 1–9.

[50] Alexander J Casson. “Wearable EEG and beyond”. In: *Biomedical engineering letters* 9.1 (2019), pp. 53–71.

[51] Johnson Thie, Alexander Klistorner, and Stuart L Graham. “Biomedical signal acquisition with streaming wireless communication for recording evoked potentials”. In: *Documenta Ophthalmologica* 125 (2012), pp. 149–159.

[52] Geetanjali Rathee, Chaker Abdelaziz Kerrache, and Muhammad Bilal. “An accurate and inter-operatable fuzzy-based system using genetic and canonical correlation analysis methods in internet-of-brain things”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 31 (2023), pp. 2726–2733.

[53] Chi-Yuan Chang et al. “Evaluation of artifact subspace reconstruction for automatic artifact components removal in multi-channel EEG recordings”. In: *IEEE Transactions on Biomedical Engineering* 67.4 (2019), pp. 1114–1121.

[54] Steven J Luck. “APA handbook of research methods in psychology, Vol. 1. Foundations, planning, measures, and psychometrics”. In: American Psychological Association, 2012. Chap. Event-related potentials, pp. 523–546.

[55] John Polich. “EEG and ERP assessment of normal aging”. In: *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section* 104.3 (1997), pp. 244–256.

[56] John Polich and Kathryn L Herbst. “P300 as a clinical assay: rationale, evaluation, and findings”. In: *International Journal of Psychophysiology* 38.1 (2000), pp. 3–19.

[57] Rik van Dinteren et al. “P300 development across the lifespan: a systematic review and meta-analysis”. In: *PloS one* 9.2 (2014), e87347.

[58] Hiroshi Shibasaki and Mark Hallett. “What is the Bereitschaftspotential?” In: *Clinical neurophysiology* 117.11 (2006), pp. 2341–2356.

[59] Natalie Mrachacz-Kersting et al. “Efficient neuroplasticity induction in chronic stroke patients by an associative brain-computer interface”. In: *Journal of neurophysiology* 115.3 (2016), pp. 1410–1421.

[60] Fatemeh Karimi et al. “Detection of movement related cortical potentials from EEG using constrained ICA for brain-computer interface applications”. In: *Frontiers in neuroscience* 11 (2017), p. 356.

[61] M Zwecker et al. “Detection of cognitive impairment in multiple sclerosis based on P300 event-related potential”. In: *Int. J. Phys. Med. Rehabil* 6.04 (2018).

[62] Yu-Cheng Chang et al. “Delay Analysis in Closed-Loop EEG Phase-Triggered Transcranial Magnetic Stimulation”. In: *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE. 2023, pp. 1–4.

[63] Ali Hossaini et al. “A Functional BCI Model by the P2731 working group: Physiology”. In: *Brain-Computer Interfaces* 8.3 (2021), pp. 54–81.

[64] Alyssa Hillary Zisk et al. "P300 latency jitter and its correlates in people with amyotrophic lateral sclerosis". In: *Clinical Neurophysiology* 132.2 (2021), pp. 632–642.

[65] Miguel Angel Lopez-Gordo, Pablo Padilla, and F Pelayo Valle. "A proposal for bio-synchronized transmission of EEG/ERP data". In: *Eurasip Journal on Wireless Communications and Networking* 2016 (2016), pp. 1–12.

[66] Alejandro Ojeda et al. "SimBSI: an open-source Simulink library for developing closed-loop brain signal interfaces in animals and humans". In: *Biomedical physics & engineering express* 6.3 (2020), p. 035023.

[67] John Polich. "Updating P300: an integrative theory of P3a and P3b". In: *Clinical neurophysiology* 118.10 (2007), pp. 2128–2148.

[68] Patrick Ofner et al. "Upper limb movements can be decoded from the time-domain of low-frequency EEG". In: *PloS one* 12.8 (2017), e0182578.

[69] Arnaud Delorme and Scott Makeig. "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis". In: *Journal of neuroscience methods* 134.1 (2004), pp. 9–21.

[70] Mads Jochumsen et al. "Quantification of movement-related EEG correlates associated with motor training: A study on movement-related cortical potentials and sensorimotor rhythms". In: *Frontiers in human neuroscience* 11 (2017), p. 604.

[71] Sunghan Lee et al. "Simultaneous EEG acquisition system for multiple users: development and related issues". In: *Sensors* 19.20 (2019), p. 4592.

[72] Michael Foerster et al. "A synchronization method for wireless acquisition systems, application to brain computer interfaces". In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2013, pp. 830–833.

[73] Karim Bennys et al. "Diagnostic value of event-related evoked potentials N200 and P300 subcomponents in early diagnosis of Alzheimer's disease and mild cognitive impairment". In: *Journal of Clinical Neurophysiology* 24.5 (2007), pp. 405–412.

[74] Pasquale Arpaia et al. "Comparing artifact removal techniques for daily-life electroencephalography with few channels". In: *2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. IEEE. 2022, pp. 1–6.

[75] Pasquale Arpaia et al. "Low-density EEG correction with multivariate decomposition and subspace reconstruction". In: *IEEE Sensors Journal* (2023).

[76] Kevin T Sweeney, Tomás E Ward, and Seán F McLoone. "Artifact removal in physiological signals—Practices and possibilities". In: *IEEE transactions on information technology in biomedicine* 16.3 (2012), pp. 488–500.

[77] Simone Corbellini et al. "Cloud based sensor network for environmental monitoring". In: *Measurement* 118 (2018), pp. 354–361.

[78] Luca Lombardo et al. "Wireless sensor network for indoor and outdoor atmospheric monitoring in culture heritage". In: *3rd IMEKO International Conference on Metrology for Archaelogy and Cultural Heritage*. 2017.

[79] William O Tatum IV. *Handbook of EEG interpretation*. Springer Publishing Company, 2021.

[80] Jose Antonio Urigüen and Begoña Garcia-Zapirain. “EEG artifact removal—state-of-the-art and guidelines”. In: *Journal of neural engineering* 12.3 (2015), p. 031001.

[81] Rakesh Ranjan, Bikash Chandra Sahana, and Ashish Kumar Bhandari. “Ocular artifact elimination from electroencephalography signals: A systematic review”. In: *Biocybernetics and Biomedical Engineering* 41.3 (2021), pp. 960–996.

[82] William O Tatum, Barbara A Dworetzky, and Donald L Schomer. “Artifact and recording concepts in EEG”. In: *Journal of clinical neurophysiology* 28.3 (2011), pp. 252–263.

[83] Ke Zeng and Xiaoli Li. “Artifact removal in EEG recordings”. In: *Signal Processing in Neuroscience* (2016), pp. 77–98.

[84] Rodney J Croft and Robert J Barry. “EOG correction: Which regression should we use?” In: *Psychophysiology* 37.1 (2000), pp. 123–125.

[85] Christian Andreas Edgar Kothe and Tzyy-Ping Jung. *Artifact removal techniques with signal reconstruction.* US Patent App. 14/895,440. 2016.

[86] Sarah Blum et al. “A Riemannian modification of artifact subspace reconstruction for EEG artifact handling”. In: *Frontiers in human neuroscience* 13 (2019), p. 141.

[87] Rasmus Bro and Age K Smilde. “Principal component analysis”. In: *Analytical methods* 6.9 (2014), pp. 2812–2831.

[88] Te-Won Lee and Te-Won Lee. *Independent component analysis.* Springer, 1998.

[89] Bruce Thompson. “Canonical correlation analysis.” In: (2000).

[90] Michael Unser and Akram Aldroubi. “A review of wavelets in biomedical applications”. In: *Proceedings of the IEEE* 84.4 (1996), pp. 626–638.

[91] Hualou Liang et al. “Empirical mode decomposition: a method for analyzing neural data”. In: *Neurocomputing* 65 (2005), pp. 801–807.

[92] Dmytro Iatsenko, Peter VE McClintock, and Aneta Stefanovska. “Nonlinear mode decomposition: A noise-robust, adaptive decomposition method”. In: *Physical Review E* 92.3 (2015), p. 032916.

[93] Jesus Minguillon, M Angel Lopez-Gordo, and Francisco Pelayo. “Trends in EEG-BCI for daily-life: Requirements for artifact removal”. In: *Biomedical Signal Processing and Control* 31 (2017), pp. 407–418.

[94] Wajid Mumtaz, Suleman Rasheed, and Alina Irfan. “Review of challenges associated with the EEG artifact removal methods”. In: *Biomedical Signal Processing and Control* 68 (2021), p. 102741.

[95] Malik Muhammad Naeem Mannan, Muhammad Ahmad Kamran, and Myung Yung Jeong. “Identification and removal of physiological artifacts from electroencephalogram signals: A review”. In: *Ieee Access* 6 (2018), pp. 30630–30652.

[96] Stefan Ehrlich. `https://github.com/stefan-ehrlich/dataset-automaticArtifactRemoval`. Last access: 2022-08.

[97] *Introducing the actiCHamp Plus – offering active and passive electrode recordings . . . and more!* `https://pressrelease.brainproducts.com/actichamp_plus/`. Last access: 2022-08.

[98] Abdel-Ouahab Boudraa and Jean-Christophe Cexus. "EMD-based signal filtering". In: *IEEE transactions on instrumentation and measurement* 56.6 (2007), pp. 2196–2202.

[99] Ali Komaty et al. "EMD-based filtering using similarity measure between probability density functions of IMFs". In: *IEEE Transactions on Instrumentation and Measurement* 63.1 (2013), pp. 27–34.

[100] Zhaohua Wu and Norden E Huang. "Ensemble empirical mode decomposition: a noise-assisted data analysis method". In: *Advances in adaptive data analysis* 1.01 (2009), pp. 1–41.

[101] Ke Zeng et al. "An EEMD-ICA approach to enhancing artifact rejection for noisy multivariate neural data". In: *IEEE transactions on neural systems and rehabilitation engineering* 24.6 (2015), pp. 630–638.

[102] Xun Chen et al. "Removing muscle artifacts from EEG data: Multichannel or single-channel techniques?" In: *IEEE Sensors Journal* 16.7 (2015), pp. 1986–1997.

[103] Gang Wang et al. "The removal of EOG artifacts from EEG signals using independent component analysis and multivariate empirical mode decomposition". In: *IEEE journal of biomedical and health informatics* 20.5 (2015), pp. 1301–1308.

[104] Xueyuan Xu, Aiping Liu, and Xun Chen. "A novel few-channel strategy for removing muscle artifacts from multichannel EEG data". In: *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE. 2017, pp. 976–980.

[105] Xun Chen et al. "The use of multivariate EMD and CCA for denoising muscle artifacts from few-channel EEG recordings". In: *IEEE transactions on instrumentation and measurement* 67.2 (2017), pp. 359–370.

[106] Yan Liu et al. "An efficient and robust muscle artifact removal method for few-channel EEG". In: *IEEE Access* 7 (2019), pp. 176036–176050.

[107] Xun Chen et al. "A novel EEMD-CCA approach to removing muscle artifacts for pervasive EEG". In: *IEEE Sensors Journal* 19.19 (2018), pp. 8420–8431.

[108] Ajay Kumar Maddirala and Rafi Ahamed Shaik. "Removal of EOG artifacts from single channel EEG signals using combined singular spectrum analysis and adaptive noise canceler". In: *IEEE Sensors Journal* 16.23 (2016), pp. 8279–8287.

[109] Rajesh Patel et al. "Suppression of eye-blink associated artifact using single channel EEG data by combining cross-correlation with empirical mode decomposition". In: *IEEE Sensors Journal* 16.18 (2016), pp. 6947–6954.

[110] Christos Stergiadis, Vasiliki-Despoina Kostaridou, and Manousos A Klados. "Which BSS method separates better the EEG Signals? A comparison of five different algorithms". In: *Biomedical Signal Processing and Control* 72 (2022), p. 103292.

[111] Tim R Mullen et al. "Real-time neuroimaging and cognitive monitoring using wearable dry EEG". In: *IEEE transactions on biomedical engineering* 62.11 (2015), pp. 2553–2567.

[112] Andrea Cataldo et al. "A Method for Optimizing the Artifact Subspace Reconstruction Performance in Low-Density EEG". In: *IEEE Sensors Journal* 22.21 (2022), pp. 21257–21265.

[113] Danilo P Mandic et al. “Empirical mode decomposition-based time-frequency analysis of multivariate signals: The power of adaptive data analysis”. In: *IEEE signal processing magazine* 30.6 (2013), pp. 74–86.

[114] Yi Zhang et al. “Noise-assisted multivariate empirical mode decomposition for multichannel EMG signals”. In: *Biomedical engineering online* 16.1 (2017), pp. 1–17.

[115] Alessandra Anzolin et al. “SEED-G: simulated EEG data generator for testing connectivity algorithms”. In: *Sensors* 21.11 (2021), p. 3632.

[116] *EEGdenoiseNet*. `https://github.com/ncclabsustech/EEGdenoiseNet`. Last access: 2022-08.

[117] Matteo Dora and David Holcman. “Adaptive single-channel EEG artifact removal with applications to clinical monitoring”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 30 (2022), pp. 286–295.

[118] Velu Prabhakar Kumaravel et al. “Efficient artifact removal from low-density wearable EEG using artifacts subspace reconstruction”. In: *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE. 2021, pp. 333–336.

[119] Quan Liu et al. “EEG artifacts reduction by multivariate empirical mode decomposition and multiscale entropy for monitoring depth of anaesthesia during surgery”. In: *Medical & biological engineering & computing* 55 (2017), pp. 1435–1450.

[120] Patrick E McKnight and Julius Najab. “Mann-Whitney U Test”. In: *The Corsini encyclopedia of psychology* (2010), pp. 1–1.

[121] Matthew J Page et al. “Updating guidance for reporting systematic reviews: development of the PRISMA 2020 statement”. In: *Journal of clinical epidemiology* 134 (2021), pp. 103–112.

[122] *VIM: International Vocabulary of Metrology*. `https://www.bipm.org/en/committees/jc/jcgm/publications`.

[123] *Evaluation of measurement data — Guide to the expression of uncertainty in measurement*. `https://www.bipm.org/en/committees/jc/jcgm/publications`.

[124] Clemens Brunner et al. “BCI Competition 2008–Graz data set A”. In: *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology* 16 (2008), pp. 1–6.

[125] Robert Leeb et al. “Brain–computer communication: motivation, aim, and impact of exploring a virtual apartment”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 15.4 (2007), pp. 473–482.

[126] Benjamin Blankertz et al. “The BCI competition III: Validating alternative approaches to actual BCI problems”. In: *IEEE transactions on neural systems and rehabilitation engineering* 14.2 (2006), pp. 153–159.

[127] Michael Tangermann et al. “Review of the BCI competition IV”. In: *Frontiers in neuroscience* 6 (2012), p. 55.

[128] Hohyun Cho et al. “EEG datasets for motor imagery brain–computer interface”. In: *GigaScience* 6.7 (2017), gix034.

[129] Robin Tibor Schirrmeister et al. “Deep learning with convolutional neural networks for EEG decoding and visualization”. In: *Human brain mapping* 38.11 (2017), pp. 5391–5420.

[130] Gerwin Schalk et al. “BCI2000: a general-purpose brain-computer interface (BCI) system”. In: *IEEE Transactions on biomedical engineering* 51.6 (2004), pp. 1034–1043.

[131] Spiros Nikolopoulos et al. “A multimodal dataset for authoring and editing multimedia content: The MAMEM project”. In: *Data in brief* 15 (2017), pp. 1048–1056.

[132] David Steyrl et al. “Motor imagery brain-computer interfaces: random forests vs regularized LDA-non-linear beats linear”. In: *Proceedings of the 6th International Brain-Computer Interface Conference*. 2014, pp. 241–244.

[133] Rongnian Tang, Zibo Li, and Xiaofeng Xie. “Motor imagery EEG signal classification using upper triangle filter bank auto-encode method”. In: *Biomedical Signal Processing and Control* 68 (2021), p. 102608.

[134] Xiaofeng Xie et al. “Motor imagery classification based on bilinear sub-manifold learning of symmetric positive-definite matrices”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25.6 (2016), pp. 504–516.

[135] Biao Sun et al. “EEG motor imagery classification with sparse spectrotemporal decomposition and deep learning”. In: *IEEE Transactions on Automation Science and Engineering* 18.2 (2020), pp. 541–551.

[136] Victoria Peterson et al. “A penalized time-frequency band feature selection and classification procedure for improved motor intention decoding in multichannel EEG”. In: *Journal of neural engineering* 16.1 (2019), p. 016019.

[137] Brenda E Olivas-Padilla and Mario I Chacon-Murguia. “Classification of multiple motor imagery using deep convolutional neural networks and spatial filters”. In: *Applied Soft Computing* 75 (2019), pp. 461–472.

[138] Qingsong Ai et al. “Feature extraction of four-class motor imagery EEG signals based on functional brain network”. In: *Journal of neural engineering* 16.2 (2019), p. 026032.

[139] Lei Sun et al. “An advanced bispectrum features for EEG-based motor imagery classification”. In: *Expert Systems with Applications* 131 (2019), pp. 9–19.

[140] Xianlun Tang et al. “Motor imagery EEG recognition based on conditional optimization empirical mode decomposition and multi-scale convolutional neural network”. In: *Expert Systems with Applications* 149 (2020), p. 113285.

[141] Minmin Miao et al. “An adaptive multi-domain feature joint optimization framework based on composite kernels and ant colony optimization for motor imagery EEG classification”. In: *Biomedical Signal Processing and Control* 61 (2020), p. 101994.

[142] Muhammad Hamza Bhatti et al. “Soft computing-based EEG classification by optimal feature selection and neural networks”. In: *IEEE Transactions on Industrial Informatics* 15.10 (2019), pp. 5747–5754.

[143] Ramesh Srinivasan et al. “Spatial sampling and filtering of EEG with spline laplacians to estimate cortical potentials”. In: *Brain topography* 8.4 (1996), pp. 355–366.

[144] Yong Zhang, Xiaomin Ji, and Yuting Zhang. “Classification of EEG signals based on AR model and approximate entropy”. In: *2015 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2015, pp. 1–6.

[145] L Xu et al. “Motor execution and motor imagery: a comparison of functional connectivity patterns based on graph theory”. In: *Neuroscience* 261 (2014), pp. 184–194.

[146] Jiaxin Xie et al. “Combining generative adversarial networks and multi-output CNN for motor imagery classification”. In: *Journal of neural engineering* 18.4 (2021), p. 046026.

[147] M Serdar Bascil, Ahmet Y Tesneli, and Feyzullah Temurtas. “Spectral feature extraction of EEG signals and pattern recognition during mental tasks of 2-D cursor movements for BCI using SVM and ANN”. In: *Australasian physical & engineering sciences in medicine* 39 (2016), pp. 665–676.

[148] M Serdar Bascil, Ahmet Y Tesneli, and Feyzullah Temurtas. “Multi-channel EEG signal feature extraction and pattern recognition on horizontal mental imagination task of 1-D cursor movement for brain computer interface”. In: *Australasian physical & engineering sciences in medicine* 38 (2015), pp. 229–239.

[149] Yousef Rezaei Tabar and Ugur Halici. “A novel deep learning approach for classification of EEG motor imagery signals”. In: *Journal of neural engineering* 14.1 (2016), p. 016003.

[150] Hyeon Kyu Lee and Young-Seok Choi. “Application of continuous wavelet transform and convolutional neural network in decoding motor imagery brain-computer interface”. In: *Entropy* 21.12 (2019), p. 1199.

[151] Kai Keng Ang et al. “Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b”. In: *Frontiers in neuroscience* 6 (2012), p. 39.

[152] Leo Breiman. *Classification and regression trees*. Routledge, 2017.

[153] Marti A. Hearst et al. “Support vector machines”. In: *IEEE Intelligent Systems and their applications* 13.4 (1998), pp. 18–28.

[154] Md Khademul Islam Molla et al. “Trial regeneration with subband signals for motor imagery classification in BCI paradigm”. In: *IEEE Access* 9 (2021), pp. 7632–7642.

[155] Shalu Chaudhary et al. “Convolutional neural network based approach towards motor imagery tasks EEG signals classification”. In: *IEEE Sensors Journal* 19.12 (2019), pp. 4494–4500.

[156] Hongtao Wang et al. “An approach of one-vs-rest filter bank common spatial pattern and spiking neural networks for multiple motor imagery decoding”. In: *Ieee Access* 8 (2020), pp. 86850–86861.

[157] Ali Al-Saegh, Shefa A Dawwd, and Jassim M Abdul-Jabbar. “Deep learning for motor imagery EEG-based classification: A review”. In: *Biomedical Signal Processing and Control* 63 (2021), p. 102172.

[158] Yannick Roy et al. “Deep learning-based electroencephalography analysis: a systematic review”. In: *Journal of neural engineering* 16.5 (2019), p. 051001.

[159] Md Zahangir Alom et al. “The history began from alexnet: A comprehensive survey on deep learning approaches”. In: *arXiv preprint arXiv:1803.01164* (2018).

[160] Md Zahangir Alom et al. “A state-of-the-art survey on deep learning theory and architectures”. In: *electronics* 8.3 (2019), p. 292.

[161] Vivienne Sze et al. “Efficient processing of deep neural networks: A tutorial and survey”. In: *Proceedings of the IEEE* 105.12 (2017), pp. 2295–2329.

[162] Alois Schlögl et al. “Evaluation criteria in BCI research”. In: *Toward Brain-Computer Interfacing*. MIT Press, 2007, pp. 327–342.

[163] Chen-Chen Fan et al. “Bilinear neural network with 3-D attention for brain decoding of motor imagery movements from the human EEG”. In: *Cognitive Neurodynamics* 15 (2021), pp. 181–189.

[164] Youngjoo Kim et al. “Correlation assisted strong uncorrelating transform complex common spatial patterns for spatially distant channel data”. In: *Computational Intelligence and Neuroscience* 2018.1 (2018), p. 4281230.

[165] Benjamin Blankertz et al. “Optimizing spatial filters for robust EEG single-trial analysis”. In: *IEEE Signal processing magazine* 25.1 (2007), pp. 41–56.

[166] John Shawe-Taylor and Shiliang Sun. “A review of optimization methodologies in support vector machines”. In: *Neurocomputing* 74.17 (2011), pp. 3609–3618.

[167] Nitesh Singh Malan and Shiru Sharma. “Time window and frequency band optimization using regularized neighbourhood component analysis for Multi-View Motor Imagery EEG classification”. In: *Biomedical Signal Processing and Control* 67 (2021), p. 102550.

[168] Jing Luo et al. “Motor imagery EEG classification based on ensemble support vector learning”. In: *Computer methods and programs in biomedicine* 193 (2020), p. 105464.

[169] Robert Tibshirani. “Regression shrinkage and selection via the lasso”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 58.1 (1996), pp. 267–288.

[170] Ruilong Zhang et al. “A novel hybrid deep learning scheme for four-class motor imagery classification”. In: *Journal of neural engineering* 16.6 (2019), p. 066004.

[171] Lie Yang et al. “A novel motor imagery EEG decoding method based on feature separation”. In: *Journal of Neural Engineering* 18.3 (2021), p. 036022.

[172] Roozbeh Zarei et al. “A PCA aided cross-covariance scheme for discriminative feature extraction from EEG signals”. In: *Computer methods and programs in biomedicine* 146 (2017), pp. 47–57.

[173] Yangyang Miao et al. “Learning Common Time-Frequency-Spatial Patterns for Motor Imagery Classification”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 29 (2021), pp. 699–707.

[174] Mamunur Rashid et al. “The classification of motor imagery response: an accuracy enhancement through the ensemble of random subspace k-NN”. In: *PeerJ Computer Science* 7 (2021), e374.

[175] Fei Wang et al. “Motor imagery classification using geodesic filtering common spatial pattern and filter-bank feature weighted support vector machine”. In: *Review of Scientific Instruments* 91.3 (2020), p. 034106.

[176] Vasilisa Mishuhina and Xudong Jiang. “Complex common spatial patterns on time-frequency decomposed EEG for brain-computer interface”. In: *Pattern Recognition* 115 (2021), p. 107918.

[177] Jing Jiang et al. “Temporal combination pattern optimization based on feature selection method for motor imagery BCIs”. In: *Frontiers in Human Neuroscience* 14 (2020), p. 231.

[178] Diego Collazos-Huertas et al. “Enhanced multiple instance representation using time-frequency atoms in motor imagery classification”. In: *Frontiers in neuroscience* 14 (2020), p. 155.

[179] Mohammad Mahdi Togha, Mohammad Reza Salehi, and Ebrahim Abiri. “An improved version of local activities estimation to enhance motor imagery classification”. In: *Biomedical Signal Processing and Control* 66 (2021), p. 102485.

[180] Moein Radman et al. “Feature fusion for improving performance of motor imagery brain-computer interface system”. In: *Biomedical Signal Processing and Control* 68 (2021), p. 102763.

[181] Jing Luo, Zuren Feng, and Na Lu. “Spatio-temporal discrepancy feature for classification of motor imageries”. In: *Biomedical Signal Processing and Control* 47 (2019), pp. 137–144.

[182] Jie Wang et al. “Feature subset and time segment selection for the classification of EEG data based motor imagery”. In: *Biomedical Signal Processing and Control* 61 (2020), p. 102026.

[183] Muhammad Tariq Sadiq et al. “Identification of motor and mental imagery EEG in two and multiclass subject-dependent tasks using successive decomposition index”. In: *Sensors* 20.18 (2020), p. 5283.

[184] Li Zhang et al. “Ensemble classifier based on optimized extreme learning machine for motor imagery classification”. In: *Journal of neural engineering* 17.2 (2020), p. 026004.

[185] Aysa Jafarifarmand et al. “A new self-regulated neuro-fuzzy framework for classification of EEG signals in motor imagery BCI”. In: *IEEE transactions on fuzzy systems* 26.3 (2017), pp. 1485–1497.

[186] Ce Zhang, Young-Keun Kim, and Azim Eskandarian. “EEG-inception: an accurate and robust end-to-end neural network for EEG-based motor imagery classification”. In: *Journal of Neural Engineering* 18.4 (2021), p. 046014.

[187] Xiuling Liu et al. “Multiscale space-time-frequency feature-guided multitask learning CNN for motor imagery EEG classification”. In: *Journal of Neural Engineering* 18.2 (2021), p. 026003.

[188] Lie Yang et al. “Motor imagery EEG decoding method based on a discriminative feature learning strategy”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 29 (2021), pp. 368–379.

[189] Ming-ai Li and Zi-wei Ruan. “A novel decoding method for motor imagery tasks with 4D data representation and 3D convolutional neural networks”. In: *Journal of Neural Engineering* 18.4 (2021), p. 046029.

[190] Juntao Xue et al. “A Multifrequency Brain Network-Based Deep Learning Framework for Motor Imagery Decoding”. In: *Neural Plasticity* 2020 (2020).

[191] Jesus L Lobo et al. “Spiking neural networks and online learning: An overview and perspectives”. In: *Neural Networks* 121 (2020), pp. 88–100.

[192] Shaveta Dargan et al. “A survey of deep learning and its applications: a new paradigm to machine learning”. In: *Archives of Computational Methods in Engineering* 27 (2020), pp. 1071–1092.

[193] Vinay Jayaram and Alexandre Barachant. “MOABB: trustworthy algorithm benchmarking for BCIs”. In: *Journal of neural engineering* 15.6 (2018), p. 066011.

[194] Andrea Apicella et al. “EEG-based measurement system for monitoring student engagement in learning 4.0”. In: *Scientific Reports* 12.1 (2022), p. 5857.

[195] Rosa Mikeal Martey et al. “Measuring game engagement: multiple methods and construct complexity”. In: *Simulation & Gaming* 45.4-5 (2014), pp. 528–547.

[196] Lina Elsherif Ismail and Waldemar Karwowski. “Applications of EEG indices for the quantification of human cognitive performance: A systematic review and bibliometric analysis”. In: *Plos one* 15.12 (2020), e0242857.

[197] M Dewan, Mahbub Murshed, and Fuhua Lin. “Engagement detection in online learning: a review”. In: *Smart Learning Environments* 6.1 (2019), pp. 1–20.

[198] Asma Ben Khedher, Imène Jraidi, and Claude Frasson. “Tracking students’ mental engagement using EEG signals during an interaction with a virtual learning environment”. In: *Journal of Intelligent Learning Systems and Applications* 11.1 (2019), pp. 1–14.

[199] Konrad Biercewicz, Mariusz Borawski, Jarosław Duda, et al. “Method for selecting an engagement index for a specific type of game using cognitive neuroscience”. In: *International Journal of Computer Games Technology* 2020 (2020).

[200] Caroline Elizabeth Brett, Catherine Sykes, and Renata Pires-Yfantouda. “Interventions to increase engagement with rehabilitation in adults with acquired brain injury: a systematic review”. In: *Neuropsychological rehabilitation* 27.6 (2017), pp. 959–982.

[201] Geoffrey Hookham and Keith Nesbitt. “A systematic review of the definition and measurement of engagement in serious games”. In: *Proceedings of the australasian computer science week multiconference.* 2019, pp. 1–10.

[202] Elena Di Lascio, Shkurta Gashi, and Silvia Santini. “Unobtrusive assessment of students’ emotional engagement during lectures using electrodermal activity sensors”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2.3 (2018), pp. 1–21.

[203] Diana K Darnell and Paul A Krieg. “Student engagement, assessed using heart rate, shows no reset following active learning sessions in lectures”. In: *PloS one* 14.12 (2019), e0225709.

[204] JK Nuamah and Younho Seong. “Support vector machine (SVM) classification of cognitive tasks based on electroencephalography (EEG) engagement index”. In: *Brain-Computer Interfaces* 5.1 (2018), pp. 1–12.

[205] Nataliya Kosmyna and Pattie Maes. “AttentivU: an EEG-based closed-loop biofeedback system for real-time monitoring and improvement of engagement for personalized learning”. In: *Sensors* 19.23 (2019), p. 5200.

[206] Alan T Pope, Edward H Bogart, and Debbie S Bartolome. “Biocybernetic system evaluates indices of operator engagement in automated task”. In: *Biological psychology* 40.1-2 (1995), pp. 187–195.

[207] Frédéric Dehais et al. “Monitoring pilot’s cognitive fatigue with engagement features in simulated and actual flight conditions using an hybrid fNIRS-EEG passive BCI”. In: *2018 IEEE international conference on systems, man, and cybernetics (SMC)*. IEEE. 2018, pp. 544–549.

[208] Rodika Sokoliuk et al. “Two spatially distinct posterior alpha sources fulfill different functional roles in attention”. In: *Journal of Neuroscience* 39.36 (2019), pp. 7183–7194.

[209] Antoine Gaume, Gérard Dreyfus, and François-Benoıˆt Vialatte. “A cognitive brain–computer interface monitoring sustained attentional variations during a continuous task”. In: *Cognitive neurodynamics* 13 (2019), pp. 257–269.

[210] Colin M MacLeod. “The Stroop task: The" gold standard" of attentional measures.” In: *Journal of Experimental Psychology: General* 121.1 (1992), p. 12.

[211] Mirko Schmidt, Valentin Benzing, and Mario Kamer. “Classroom-based physical activity breaks and children’s attention: Cognitive engagement works!” In: *Frontiers in psychology* 7 (2016), p. 208144.

[212] Marsha E Bates and Edward P Lemay. “The d2 Test of attention: construct validity and extensions in scoring techniques”. In: *Journal of the International Neuropsychological Society* 10.3 (2004), pp. 392–400.

[213] Konrad Kulikowski and Jarosław Orzechowski. “Working memory and fluid intelligence as predictors of work engagement—Testing preliminary models”. In: *Applied Cognitive Psychology* 33.4 (2019), pp. 596–616.

[214] Laurens R Krol, Sarah-Christin Freytag, and Thorsten O Zander. “Meyendtris: A hands-free, multimodal tetris clone using eye tracking and passive BCI for intuitive neuroadaptive gaming”. In: *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. 2017, pp. 433–437.

[215] Ankur Joshi et al. “Likert scale: Explored and explained”. In: *British journal of applied science & technology* 7.4 (2015), pp. 396–403.

[216] George H Klem. “The ten-twenty electrode system of the international federation. The international federation of clinical neurophysiology”. In: *Electroencephalogr. Clin. Neurophysiol. Suppl.* 52 (1999), pp. 3–6.

[217] Wen-Sheng Chang et al. “The association between working memory precision and the nonlinear dynamics of frontal and parieto-occipital EEG activity”. In: *Scientific reports* 13.1 (2023), p. 14252.

[218] Andrea Apicella et al. “A survey on eeg-based solutions for emotion recognition with a low number of channels”. In: *IEEE Access* 10 (2022), pp. 117411–117428.

[219] Feng Li et al. “Deep models for engagement assessment with scarce label information”. In: *IEEE Transactions on Human-Machine Systems* 47.4 (2016), pp. 598–605.

[220] Paul Sauseng et al. “Fronto-parietal EEG coherence in theta and upper alpha reflect central executive functions of working memory”. In: *International journal of Psychophysiology* 57.2 (2005), pp. 97–103.

[221] Michael B Steinborn et al. “Methodology of performance scoring in the d2 sustained-attention test: Cumulative-reliability functions and practical guidelines.” In: *Psychological assessment* 30.3 (2018), p. 339.

[222] Purya Baghaei, Hamdollah Ravand, and Mahsa Nadri. “Is the d2 test of attention Rasch scalable? Analysis with the Rasch Poisson counts model”. In: *Perceptual and Motor Skills* 126.1 (2019), pp. 70–86.

[223] Bee Wah Yap and Chiaw Hock Sim. “Comparisons of various types of normality tests”. In: *Journal of Statistical Computation and Simulation* 81.12 (2011), pp. 2141–2155.

[224] Almudena Capilla et al. “The natural frequencies of the resting human brain: An MEG-based atlas”. In: *NeuroImage* 258 (2022), p. 119373.

[225] Dominik Welke and Edward A Vessel. “Naturalistic viewing conditions can increase task engagement and aesthetic preference but have only minimal impact on EEG quality”. In: *NeuroImage* 256 (2022), p. 119218.

[226] Boyang Xu et al. “A Hybrid BCI System Integrating Motor Imagery and SSVEP for Wheelchair Control”. In: *Proceedings of the 2024 13th International Conference on Computing and Pattern Recognition.* 2024, pp. 353–360.

[227] Yuyi Lu et al. “Feature Extraction and Classification of Motor Imagery EEG Signals in Motor Imagery for Sustainable Brain–Computer Interfaces”. In: *Sustainability* 16.15 (2024), p. 6627.

[228] Pasquale Arpaia et al. “Visual and haptic feedback in detecting motor imagery within a wearable brain–computer interface”. In: *Measurement* 206 (2023), p. 112304.

[229] Huanyu Wu, Siyang Li, and Dongrui Wu. “Motor imagery classification for asynchronous EEG-based brain-computer interfaces”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* (2024).

[230] Xinzhi Ma et al. “An attention-based motor imagery brain–computer interface system for lower limb exoskeletons”. In: *Review of Scientific Instruments* 95.12 (2024).

[231] Bilal Alchalabi, Jocelyn Faubert, and David R Labbe. “A multi-modal modified feedback self-paced BCI to control the gait of an avatar”. In: *Journal of Neural Engineering* 18.5 (2021), p. 056005.

[232] Liuyin Yang and Marc M Van Hulle. “Real-time navigation in google street view® using a motor imagery-based BCI”. In: *Sensors* 23.3 (2023), p. 1704.

[233] Lukas Jaeger et al. “How the CYBATHLON competition has advanced assistive technologies”. In: *Annual Review of Control, Robotics, and Autonomous Systems* 6.1 (2023), pp. 447–476.

[234] Lea Hehenberger et al. “Long-term mutual training for the CYBATHLON BCI race with a tetraplegic pilot: a case study on inter-session transfer and intra-session adaptation”. In: *Frontiers in Human Neuroscience* 15 (2021), p. 635777.

[235] Attila Korik et al. “Competing at the Cybathlon championship for people with disabilities: long-term motor imagery brain–computer interface training of a cybathlete who has tetraplegia”. In: *Journal of NeuroEngineering and Rehabilitation* 19.1 (2022), p. 95.

[236] Junhyuk Choi et al. “Developing a motor imagery-based real-time asynchronous hybrid BCI controller for a lower-limb exoskeleton”. In: *Sensors* 20.24 (2020), p. 7309.

[237] Imran Khan Niazi et al. “Associative cued asynchronous BCI induces cortical plasticity in stroke patients”. In: *Annals of clinical and translational neurology* 9.5 (2022), pp. 722–733.

[238] Hongtao Wang et al. “The control of a virtual automatic car based on multiple patterns of motor imagery BCI”. In: *Medical & biological engineering & computing* 57 (2019), pp. 299–309.

[239] Jiaxing Wang et al. “Towards a Fast and Robust MI-BCI: Online Adaptation of Stimulus Paradigm and Classification Model”. In: *IEEE Transactions on Instrumentation and Measurement* (2024).

[240] Bonkon Koo et al. “A hybrid NIRS-EEG system for self-paced brain computer interface with online motor imagery”. In: *Journal of neuroscience methods* 244 (2015), pp. 26–32.

[241] Jeong-Hyun Cho, Ji-Hoon Jeong, and Seong-Whan Lee. “NeuroGrasp: Real-time EEG classification of high-level motor imagery tasks using a dual-stage deep learning framework”. In: *IEEE Transactions on Cybernetics* 52.12 (2021), pp. 13279–13292.

[242] Enzeng Dong et al. “A multi-modal brain–computer interface based on threshold discrimination and its application in wheelchair control”. In: *Cognitive Neurodynamics* 16.5 (2022), pp. 1123–1133.

[243] Yang Yu et al. “Self-paced operation of a wheelchair based on a hybrid brain-computer interface combining motor imagery and P300 potential”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25.12 (2017), pp. 2516–2526.

[244] Christa Neuper et al. “Imagery of motor actions: Differential effects of kinesthetic and visual–motor mode of imagery in single-trial EEG”. In: *Cognitive brain research* 25.3 (2005), pp. 668–677.

[245] Jouni Pohjalainen, Okko Räsänen, and Serdar Kadioglu. “Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits”. In: *Computer Speech & Language* 29.1 (2015), pp. 145–171.

[246] Pasquale Arpaia et al. “Online processing for motor imagery-based brain-computer interfaces relying on EEG”. In: *2023 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. IEEE. 2023, pp. 01–06.

[247] Stephen M Goldfeld and Richard E Quandt. “Some tests for homoscedasticity”. In: *Journal of the American statistical Association* 60.310 (1965), pp. 539–547.

[248] Mohammad YM Naser and Sylvia Bhattacharya. “Resting EEG state, an insight into motor imagery signal characteristics”. In: *2023 15th Biomedical Engineering International Conference (BMEiCON)*. IEEE. 2023, pp. 1–5.

[249] Vernon J Lawhern et al. “EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces”. In: *Journal of neural engineering* 15.5 (2018), p. 056013.

[250] Pasquale Arpaia et al. “Sinc-EEGNet for Improving Performance While Reducing Calibration of a Motor Imagery-Based BCI”. In: *2023 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*. IEEE. 2023, pp. 1063–1068.

[251] Pasquale Arpaia et al. “Endless Runner Game in Virtual Reality Controlled by a Self-paced Brain-Computer Interface Based on EEG and Motor Imagery”. In: *International Conference on Extended Reality*. Springer. 2024, pp. 208–225.

This Ph.D. thesis has been typeset by means of the TEX-system facilities. The typesetting engine was pdfLATEX. The document class was `toptesi`, by Claudio Beccari, with option `tipotesi =scudo`. This class is available in every up-to-date and complete TEX-system installation.