

POLITECNICO DI MILANO
Corso di Laurea Magistrale in Ingegneria Biomedica
Dipartimento di Elettronica, Informazione e Bioingegneria



POLITECNICO
MILANO 1863

Implementation of a complete SSVEP-based Brain Computer Interface system for spelling

Relatore: Prof. Anna Maria Bianchi
Correlatore: Dott. Giulia Tacchino

Tesi di Laurea Magistrale di:
Beatrice Cairo, matricola 823343

Anno Accademico 2015-2016

Abstract

Brain-Computer interfaces (BCIs) are hardware and/or software systems that allow cerebral activity alone to control computers or external devices: their main purpose is to enable severely disabled people (suffering from locked-in syndrome, i.e. complete paralysis) to interact with their surroundings, without the involvement of damaged peripheral nerves and muscles. This is usually achieved by using control signals generated from electroencephalographic (EEG) activity for either written communication or actuation purposes. Different approaches are used to develop a BCI paradigm, which are classified as evoked or invoked, based on the presence or absence of an external stimulus to trigger the needed cerebral response. Among the evoked paradigms, P300- and Steady State Visual Evoked Potentials (SSVEP)-based BCIs are the most popular. In this project a SSVEP-based BCI communication system has been implemented and tested, because of the higher communication rate (defined as Information Transfer Rate, ITR) with respect to other BCI techniques and its asynchronous interface that depends only on the user's timing. This approach takes advantage of the steady-state evoked potential response of the visual cortex to flashing visual stimuli. Therefore, multiple targets encoded using different frequencies can be implemented in an interface so that, according to which target the users fix their gaze on, the frequency content of the EEG recordings will be different. However, the design of such a speller presents some difficulties in terms of stimulus presentation (which must be accurate and stable). In literature, this issue is often solved by dedicated hardware that can implement stable flickering frequencies through LEDs, but this solution implies added costs and complex, specific designs for the necessary hardware and software. Furthermore, another issue arises: efficient target identification is the most fundamental step for user satisfaction and it requires highly discriminative features and fast and accurate classifiers that require little to no training. However, SSVEP-based BCIs often implement threshold-based classifiers that identify the highest peak of the amplitude spectrum of the EEG recording as the SSVEP frequency, which requires a user-specific threshold and is not robust to noisier recordings where more spectral peaks can exceed the threshold. As such, different classification methodologies have been explored in literature.

This project aims at developing a complete BCI system on a general-purpose com-

puter, to implement the paradigm, acquire experimental EEG recordings and analyze the SSVEP response of volunteer test subjects. The first step was the design of the BCI paradigm, using a 6-by-4 grid of blinking colored stimuli, each representing a different letter of the alphabet, with blinking frequencies between 8 and 21 Hz (low and medium SSVEP bands). The interface timing is controlled by specific software for accurate stimulation and the grid is designed for user comfort (which in turns should improve the signal to noise ratio, SNR, of the acquired data). The experimental setup was also created, according to current standards on EEG acquisition for visual evoked potentials: an electrode cap was placed on the subject and the electrodes in the parieto-occipital region were used to acquire signals generated by neuronal populations in the visual cortex. The test volunteers were asked to fill in a questionnaire about their lifestyle (to analyze a possible correlation between performance and user’s habits) and were then seated in front of the interface screen, where they were asked to observe each flashing stimulus for twenty seconds. A closed eyes test was also performed, to record and study the alpha rhythm of the subject and its influence on the task. The acquired EEG was filtered and then the three electrodes with highest SNR were selected, based on one recording with known frequency.

Afterward, the preprocessed data were used to test two different classification algorithms, in order to extract the SSVEP main frequency and compare their performances. In both cases, two classifiers in series were applied: first the spectrum was estimated for the test recording at unknown SSVEP frequency using an autoregressive (AR) model. Then, the three most likely class candidates were extracted (defined as the maximum frequency peaks higher than the noise floor and within a ± 0.4 Hz range from one of the stimulation frequencies). If only one possible candidate is selected (i.e. there was no ambiguity in the classification), the target class was the one corresponding to the chosen peak. Otherwise, in case of ambiguity where more than one class is a likely candidate, the target class is identified in the next classification step.

The first classifier tested is based on temporal features. Before computing the Pearson index for cross-correlation between a test set and a training set of known SSVEP component, both time series are spatially filtered using Canonical Correlation Analysis (CCA). This allows a synthesis of the multichannel information into a vector of lower dimensionality, while the correlation between the two series is increased. This filter is applied to two pairs of time series: first, to the test set and the training set and then to the test set and the ideal sinusoids at the frequency of the training set SSVEP stimulus (and its harmonics). Afterward, the Pearson cross-correlation index is computed between the filtered sets and an averaged coefficient is obtained for each class candidate, so that the maximum can be selected as the one associated to the target class. However, a certain variability in accuracy according to the time window selected (in length and starting point) was observed, due to phase shifts between the same frequency components in the

test set and training set.

The second method implemented is a new approach based on spectral coherence, a useful index of similarity between spectra that does not depend on phase shifts. Coherence was estimated using a bivariate AR model between the test and training set (with a single channel approach to reduce dimensionality), which is a well-tested spectral estimate for EEG recordings, and is better suited to short data windows than non-parametric estimates. The maximal coherence at the frequency of the training set SSVEP is then used as the classification feature.

Results on EEG simulations demonstrated that both these approaches have a good discrimination power, although correlation is slightly superior when no phase delay is present. On the other hand, experimental results showed an overall median accuracy of $76.5 \pm 11.5\%$ for correlation and $71 \pm 25.5\%$ for coherence, with a maximum theoretical ITR of, respectively, 20.16 bits/min and 17.67 bits/min when using four-second long data windows. In literature, 70% is often considered the minimum threshold for a BCI system in clinical applications to avoid inducing frustration in the subject and average ITR values are over 20 bits/min. As such, we can conclude that the performances of the tested classifiers are coherent with the state of the art. However, the most interesting result of this study lies in the accuracy of the classifiers frequency by frequency: temporal features are more discriminative on lower frequencies (up to 15 Hz), while the coherence-based classifier outperforms the correlation-based one for the classes at higher frequencies (higher than 15 Hz): in the future, a classifier that exploits both features could be implemented to increase accuracy. Furthermore, from the present study a strong user specificity of the system emerged: a customization of the system in terms of electrode selection was already implemented and user-specific paradigm design (in terms of flickering frequencies) could also be introduced to improve performances.

Sommario

Le Brain-Computer Interfaces (BCI) sono sistemi hardware e/o software che permettono di controllare computer o strumenti esterni tramite attività cerebrale: il loro principale scopo è quello di permettere a persone con gravi disabilità (che soffrono di sindrome locked-in, cioè paralisi completa) di interagire con l'ambiente circostante, senza il coinvolgimento di nervi periferici e muscoli danneggiati. Questo è solitamente realizzato utilizzando segnali di controllo generati dall'attività elettroencefalografica (EEG) di un soggetto al fine di consentire una possibile comunicazione scritta o attuazione. Per sviluppare un paradigma BCI è possibile utilizzare diversi approcci, classificati come evocati o invocati, basati rispettivamente sulla presenza o assenza di uno stimolo esterno che induca la risposta cerebrale necessaria. Tra i paradigmi evocati, le BCI basate su P300 e Steady State Visual Evoked Potentials (SSVEP) sono le più popolari: in questo progetto, un sistema di comunicazione basato su SSVEP è stato implementato e testato, a fronte della sua maggior velocità di comunicazione (definita come Information Transfer Rate, ITR) rispetto alle altre tecniche di BCI e alla sua interfaccia asincrona che dipende solo dalla tempistica dell'utente. Questo approccio sfrutta la risposta di potenziale evocato steady state della corteccia visiva causata da stimoli visivi lampeggianti. Pertanto, è possibile implementare un'interfaccia in cui molteplici target vengono codificati usando frequenze diverse, in modo tale che, a seconda del target su cui gli utenti posano lo sguardo, il contenuto in frequenza delle registrazioni EEG sarà differente. Tuttavia, la progettazione di uno speller di questo tipo comporta alcune difficoltà in termini di presentazione degli stimoli (che devono essere accurati e stabili). Spesso, in letteratura, questo problema è risolto utilizzando hardware dedicato che può implementare frequenze di lampeggiamento stabili tramite LED, ma questo implica costi addizionali e design costosi e specifici sia per l'hardware che per il software. Inoltre, si presenta un'altra problematica riguardante l'identificazione dei target, ovvero nel passaggio più importante per la soddisfazione dell'utente, in quanto sono richiesti sia parametri di classificazione altamente discriminativi che classificatori veloci e accurati che richiedono poco o nessun training; spesso, BCI basate su SSVEP implementano classificatori a soglia che identificano il picco più alto dello spettro in ampiezza della registrazione EEG come rappresentativo della frequenza SSVEP. Tuttavia, questo richiede per il singolo

utente una soglia specifica che potrebbe però non risultare robusta a registrazioni più rumorose, dove più picchi spettrali oltrepassano la soglia. Pertanto, in letteratura si sono esplorate diverse metodologie di classificazione.

Questo progetto si pone l'obiettivo di sviluppare un sistema BCI completo su computer general purpose, per implementare il paradigma, acquisire dati EEG sperimentali e analizzare la risposta SSVEP di volontari. Il primo passo è stato la progettazione del paradigma BCI, utilizzando una scacchiera 6x4 di stimoli lampeggianti colorati, ognuno rappresentante una diversa lettera dell'alfabeto, con frequenze di lampeggiamento tra 8 e 21 Hz (ovvero le bande SSVEP basse e medie). La temporizzazione dell'interfaccia è controllata da un software specifico per stimolazione e la scacchiera è stata progettata per il comfort dell'utente (di conseguenza migliorando il rapporto segnale/rumore, SNR, dei dati acquisiti). È stato progettato anche il setup sperimentale, in accordo agli standard attuali di acquisizione EEG di potenziali evocati visivi: una cuffia con elettrodi intessuti è stata posta sulla testa del soggetto e gli elettrodi posizionati sulla regione parieto-occipitale sono stati utilizzati per acquisire i segnali generati dalle popolazioni neurali della corteccia visiva. I volontari del test hanno compilato, su richiesta dello sperimentatore, un questionario relativo al loro stile di vita (per analizzare una possibile correlazione tra performance e abitudini dell'utente) e sono stati poi fatti sedere di fronte allo schermo dell'interfaccia, chiedendo loro di osservare ogni stimolo lampeggiante per venti secondi. È stato anche svolto un test a occhi chiusi, per registrare e studiare il ritmo alfa del soggetto e la sua influenza sul task. Il segnale EEG così acquisito è poi stato filtrato; successivamente i tre elettrodi con massimo SNR sono stati selezionati usando una registrazione con frequenza nota.

I dati preprocessati sono poi stati usati per testare due differenti algoritmi di classificazione: il primo che usa parametri di correlazione temporale e il secondo che sfrutta la coerenza spettrale per estrarre la frequenza principale dell'SSVEP. In entrambi i casi, sono stati applicati due classificatori in serie: si è dapprima stimato uno spettro per la registrazione di test a frequenza SSVEP non nota, utilizzando un modello autoregressivo (AR). Successivamente, sono state trovate le tre classi candidate con maggior probabilità di essere la classe target, definite come quelle corrispondenti ai picchi massimi di frequenza, con ampiezza maggiore rispetto al rumore di fondo e all'interno della gamma ± 0.4 Hz da una delle frequenze di stimolazione. Se un unico picco era selezionato (ovvero se non c'era ambiguità nella classificazione), la classe target risultava essere quella corrispondente al picco individuato. Altrimenti, nel caso fossero identificate più classi candidate, la classe target era selezionata al passo di classificazione successivo.

Il primo classificatore implementato e testato nel presente progetto si basa su parametri temporali. Prima di calcolare l'indice di Pearson per la cross-correlazione tra un test set e un training set di componente SSVEP nota, entrambe le serie temporali sono state filtrate spazialmente usando l'Analisi di Correlazione Canonica (CCA). Questa

metodologia permette di sintetizzare l'informazione di un approccio multicanale in un vettore di dimensionalità inferiore, mentre la correlazione tra le due serie temporali è aumentata. Questo filtro è stato applicato sia alla coppia di dati di test e di training che a quella dei dati di test e delle sinusoidi ideali alla frequenza dello stimolo SSVEP di training (e le sue armoniche). Successivamente, l'indice di cross-correlazione di Pearson è stato calcolato tra i dati filtrati ed è stato ottenuto un coefficiente medio per ogni classe candidata; la classe target era quindi identificata come quella associata al massimo valore assunto dal coefficiente medio. Tuttavia, risultava essere presente una certa variabilità di accuratezza a seconda della finestra temporale osservata (in termini di lunghezza e segmento selezionato), che potrebbe essere dovuta alla presenza di uno sfasamento tra le stesse componenti in frequenza dei dati di test e di training.

Il secondo metodo implementato è un nuovo metodo di classificazione basato sulla coerenza spettrale, un efficace indice di similarità tra spettri che non dipende dallo sfasamento tra serie temporali. La coerenza è stata stimata usando un modello AR bivariato tra dati di test e di training (con un approccio a singolo canale per ridurre la dimensionalità dei dati). Il calcolo della coerenza spettrale mediante metodi parametrici rappresenta una tecnica molto usata in letteratura per l'analisi di segnali EEG e risulta essere inoltre più efficace su brevi finestre temporali rispetto a stime non parametriche. Infine, la coerenza massima alla frequenza del SSVEP dei dati di training è stata usata come parametro di classificazione.

I risultati su segnali EEG simulati hanno dimostrato che entrambi gli approcci hanno buon potere di discriminazione, sebbene la correlazione abbia accuratezza leggermente superiore se non è presente ritardo in fase. I risultati sui dati reali (EEG acquisiti sui soggetti volontari) hanno invece dimostrato un'accuratezza complessiva mediana di $76.5 \pm 11.5\%$ per la correlazione e $71 \pm 25.5\%$ per la coerenza, con ITR teorico massimo di, rispettivamente, 20.16 bits/min e 17.67 bits/min usando finestre temporali di quattro secondi. In letteratura, la soglia minima di accuratezza è spesso considerata pari al 70% per un sistema BCI in applicazioni cliniche (al fine di evitare frustrazione nel paziente) e valori di ITR medi sono superiori a 20 bits/min. Si può concludere che le performance dei classificatori testati sono coerenti con lo stato dell'arte. Ciononostante, il risultato più interessante dello studio è nell'accuratezza dei classificatori frequenza per frequenza: i parametri di correlazione sono più discriminativi sulle basse frequenze (fino a 15 Hz), mentre quelli di coerenza hanno performance migliori sulle classi a frequenza maggiore (superiore a 15 Hz). Pertanto, un classificatore che sfrutta entrambi i parametri potrebbe essere implementato in futuro per aumentare l'accuratezza di classificazione. Inoltre, dallo studio emerge una forte specificità del sistema per il singolo utente: una personalizzazione del sistema in termini di selezione degli elettrodi è già presente e la progettazione di un paradigma utente-specifico (in termini di frequenze di lampeggiamento) potrebbe essere introdotto per aumentare le performance.

Contents

1	Introduction	1
2	State of the Art	3
2.1	BCI classification	7
2.2	Comparison of SSVEP- and P300-based BCI performances	10
2.3	SSVEP-based BCI: State of the Art	13
2.4	BCI user demographics	16
3	Material and Methods	18
3.1	BCI paradigm	18
3.2	Experimental Setup	25
3.3	EEG Simulations	30
3.4	Test Subjects Recruitment	33
3.5	Data acquisition and preprocessing	34
3.6	Approach	38
4	Implementation	43
4.1	Training	43
4.2	Features extraction and classification	44
4.2.1	Correlation	44
4.2.2	Coherence	49
5	Results	54
5.1	Simulated data	54
5.2	Experimental data	56
6	Conclusions	71
A	Canonical Correlation Analysis	80

List of Figures

2.1	Generic implementation of a BCI system. (Image from [1])	3
2.2	ITR and required training time for different BCI systems. Note that SSVEP-based BCIs have higher ITR with little to no training. (Image taken from [2])	7
2.3	Evoked vs induced and active vs passive BCI classification (image from [3])	9
2.4	Flickering stimulation and consequent SSVEP response. (Image from [4])	12
2.5	Typical setup for a SSVEP-based BCI (Image taken from [4]). Note the stimulation screen, data acquisition, signal processing and biofeedback blocks.	13
3.1	A screenshot of the blinking interface. Notice the differently colored squares on a black background.	18
3.2	SNR with respect to stimulation frequency: note the three subsystems with peak SNR. (Image from [5])	20
3.3	The approximation approach: considering a refresh rate of 60 Hz, a 10 Hz frequency (picture a) can be easily implemented by switching every three frames. On the other hand an 11 Hz frequency would require switching every 2.73 frames, which is obviously impossible. Therefore, we alternate between switching every 2 or 3 frames, by using Equation (3.1). (Image from [4])	22
3.4	Schematics of the experimental setup used during this project.	25
3.5	A view of the experimental setup. In particular, a clear image of the cap and the straps, plus the wired headbox.	27
3.6	A view of the experimental setup. Presentation projects the scenario on a CRT screen and the subject is strapped to the chair in front of the monitor.	28
3.7	A view of the experimental setup. The whole setup, including the acquisition computer and the computer for trigger generation, can be observed.	29
3.8	Time trend of the simulated signal at 13 Hz before (A) and after (B) filtering.	31
3.9	Power spectral density estimate of the simulation with signal at 13 Hz. . .	32

3.10	Butterworth seventh-order filter: magnitude and phase response. As pictured, the filter has a flat magnitude response in the band of interest, with small transition bands. Furthermore, the phase is more linear than that of other filters of similar nature (e.g. Chebyshev I and II).	35
3.11	Time trend of an EEG recording at 8 Hz before (A) and after (B) filtering. The second harmonics at 16 Hz is strongly noticeable.	37
3.12	Workflow of CCA-based classification, as used in the project. (Image from [6])	40
4.1	Effect of CCA: overlapped real EEG and ideal sinusoid at same frequency before (A) and after (B) CCA. Note that similar frequency content allows for a good overlap of the signals, therefore their cross-correlation at zero time lag is increased.	47
4.2	Effect of CCA: overlapped real test EEG and train EEG at same frequency before (A) and after (B) CCA. Note that similar frequency content allows for a good overlap of the signals, therefore their cross-correlation at zero time lag is increased.	48
4.3	Spectra of the test (A) and train (B) recordings, estimated with AR model. The stimulus frequency is 8 Hz for both: note the peak at its second harmonics, 16 Hz.	52
4.4	Cross-spectrum (A) and coherence (B) between the train and test recording estimated with AR model. The stimulus frequency is 8 Hz for both: note the peak at its second harmonics, 16 Hz.	53
5.1	Questionnaire results on psychoactive substance ingestion by the test subjects.	58
5.2	Questionnaire results on sleeping habits of the test subjects (A) and their estimate of weekly hours spent on the computer (B).	58
5.3	Classification accuracy for the correlation approach across ten subjects with relation to number of electrodes used (using four-second long time windows). Notice the peak at 3 electrodes and the lack of significant improvement with higher number of electrodes.	62
5.4	Classification accuracy for the correlation (three electrodes) and coherence (one electrode) approaches across ten subjects with relation to the window length used. Notice the lack of improvement for windows longer than 4 seconds.	63
5.5	Classification accuracy for the correlation approach across ten subjects per single frequency. Note the decrease in accuracy at higher frequencies (higher than 15 Hz).	65

5.6	Classification accuracy for the coherence approach across ten subjects per single frequency. Note the higher accuracy at 17 and 19 Hz with respect to the correlation approach.	65
5.7	The overlapped recording of a subject under 8 Hz flashing stimulus and an sinusoid with frequency equal to 16 Hz. Note the strong second harmonic component of the SSVEP signal, validating its use in the classification algorithm.	67
5.8	Welch estimate of 15 Hz SSVEP recording (A) and 21 Hz SSVEP (B) on the same subject across the eight electrodes recorded. Note the difference between the two spectra in terms of SNR.	67

List of Tables

5.1	Normalized confusion matrix of the correlation approach computed across ten simulations: the row indicates target class; the column is the assigned class. Total accuracy is $95.2\pm 5.6\%$	55
5.2	Normalized confusion matrix of the coherence approach computed across ten simulations: the row indicates target class; the column is the assigned class. Total accuracy is $85.7\pm 7.3\%$	55
5.3	Confusion matrix of the correlation approach computed for one subject (male, age 23): the row indicates target class; the column is the assigned class. Total accuracy 79.17%.	60
5.4	Confusion matrix of the coherence approach computed for one subject (male, age 23): the row indicates target class; the column is the assigned class. Total accuracy 75%.	60
5.5	Percentage median accuracy and interquartile ranges for the correlation approach with respect to the number of electrodes used (using four-second long time windows).	62
5.6	Percentage median accuracies and interquartile ranges for the correlation (three electrodes) and coherence (one electrode) approaches with respect to the data length used.	63
5.7	Percentage median accuracy and interquartile ranges for the correlation approach with respect to the target frequency.	65
5.8	Percentage median accuracy and interquartile ranges for the coherence approach with respect to the target frequency.	65
5.9	Specificity and FP rate per frequency, to evaluate the nature of the classification errors.	68
5.10	Accuracy and specificity frequency by frequency across trials for the same subject in case of correlation-based classification. Note the similar misclassification rates for the higher frequencies.	70

5.11 Accuracy and specificity frequency by frequency across trials for the same subject in case of coherence-based classification. Note the low specificity at 15 Hz in both trials (i.e. the tendency of the classifier to incorrectly classify recordings as 15 Hz). 70

Chapter 1

Introduction

Locked-in syndrome (LIS) is defined as a neurological condition, in which a ventral pontine lesion causes the patients to suffer from quadriplegia and anarthria (i.e., inability to move or speak) [7]. Many disorders can disrupt the physiological neuromuscular communication, among which are amyotrophic lateral sclerosis, brainstem stroke, brain or spinal cord injury, cerebral palsy, muscular dystrophies, multiple sclerosis and others. 1% of all stroke survivors are affected by incomplete or total LIS and there are two million LIS patients in the United States alone [8]. However, these patients are conscious and aware, their survival rate is as high as 80% ten years after the insult to the ventral pons and their EEG activity is generally normal [9]. It is therefore critical to improve the quality of life of these individuals, through assistive devices.

Unlike patients in vegetative state, people suffering from LIS are usually able to communicate by means of vertical gaze, and/or upper eyelid movements: communication with eye movement tracking devices has therefore been widely researched for the past twenty years [10]. However, in some cases palsy may extend to ocular motility, making the patient unable to communicate effectively through eye movements.

In this context, brain-computer interfaces (BCIs) are being studied to make it easier for LIS patients to communicate and exert their volition: a BCI system can send motor commands to a robotic arm or allow the users to spell their thoughts, as it is controlled by brain activity, bypassing the peripheral nervous system and the musculoskeletal system [11]. These systems also appear both as a viable mode of communication for people with severe neuromuscular disorders like spinal cord injury or cerebral palsy [1] and also as neurofeedback tools for rehabilitation in patients affected by stroke, autism, ADHD and other disorders [12].

Several types of biomedical signals and images can be used to detect the cerebral activity of interest: in particular the electroencephalographic (EEG) signal can be used thanks to ad hoc classification algorithms. In fact, there are many features that can be extracted from the EEG, among which steady state visual evoke potentials (SSVEP)

and the P300 component of event related potentials (ERPs), both of which are used in BCIs, each with their own advantages and disadvantages [2, 13].

Specifically, in this project the SSVEP-based BCI paradigm was studied for communication purposes, because of its high Information Transfer Rate (ITR) and low training time. Particular attention was given to the methodologies developed by Nakanishi, M. et al in their 2014 study *A high-speed brain speller using steady-state visual evoked potentials*, implementing the stimulation and classification techniques therein adopted [6].

This report will first introduce the state of the art on BCIs, both in terms of applications, challenges and system classification, and compare the SSVEP-based and P300-based paradigms. The state of the art on SSVEP-based BCIs in terms of paradigm and classification algorithms is also described, for a better comparison of the performances found in literature and the system implemented during the course of this project. Additionally, a brief discussion on BCI users demographics will be introduced to contextualize the statistical analyses of this project, performed on the data collected from the experiment volunteers.

The system developed will be consequently analyzed in details, focusing on the paradigm design created in Neurobehavioral Systems' Presentation, the creation of the experimental setup and the development of the training and classification algorithms, written using Matlab 2016a and based on the methods developed by Nakanishi et al [6]. There will be a particular focus on the stimulation interface design and its implementation with an approximated frequency method. Furthermore, the stability and accuracy of such stimuli, administered using a personal computer, will be discussed. The acquisition of the training and test sets, the preprocessing of the EEG signals and the choice of the reference signals and testing electrodes will also be explained. Classification through cross-correlation as evaluated from the testing, training and reference signals will be analyzed in depth, in particular the spatial filters implemented through Canonical Correlation Analysis (CCA). This approach will be compared to a new methodology developed in this project based on the spectral coherence between a training set of known stimulation frequency and a test set, using autoregressive spectral estimates of the EEG data.

Lastly, the results obtained from the recordings acquired during the project will be analyzed and compared to the state of the art, conclusions on the efficacy and performance of the system created will be drawn, with potential future developments for a more efficient online classifier and the implementation of a phase classification system.

Chapter 2

State of the Art

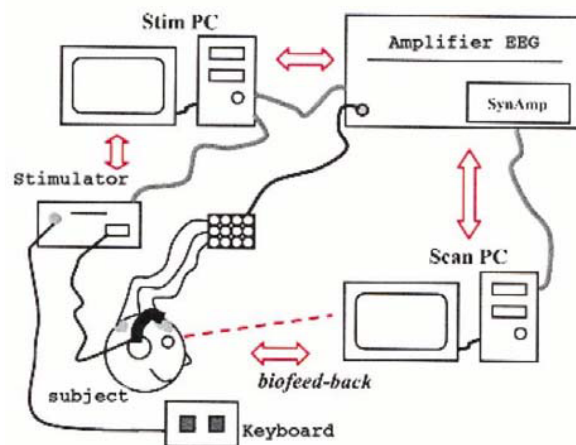


Figure 2.1: Generic implementation of a BCI system. (Image from [1])

A Brain-Computer Interface (BCI) is defined as a communication or control system in which the user's messages or commands do not depend on the brain's normal output channels (see Figure 2.1) [1]. In fact, when the central nervous system activity is physiological but the patient is not able to communicate efficiently because of a pathology of the peripheral nervous system and/or the musculoskeletal system, a BCI can bypass the damaged pathways in order to either communicate or, through robotic supports, act on the surrounding environment.

Successful BCI applications, especially in "real-world" settings, depend on the ability of the users to voluntarily modulate their neural activity, producing signals that the classification software can easily and accurately detect, even when background mental activity and electrical noise produce a disturbance in the recording. As such, it is important that the users are trained in the use of the system, through repeated practice

and adaptive feedback systems optimized for the specific end-users.

In this framework, it can be said that the fundamental components of a BCI include a signal acquisition system, an analogic front-end for amplification, sampling and quantization of the analog signal, software for feature extraction and classification, an output (device commands) and a protocol to guide its operation [14]. Therefore, during the implementation of a system, each of these steps has to be carefully evaluated and designed.

Concerning the acquisition system, various techniques are used in BCI to monitor brain activity and specifically to acquire the signal of interest: neuroimaging like functional magnetic resonance imaging (fMRI) and functional near-infrared spectroscopy (fNIRS); invasive and non-invasive recording of the neural electrical activity, like EEG, electrocorticogram (ECoG), magnetoencephalography (MEG) and neuronal action potentials recording [15]. Among them, EEG acquisition systems with non-implanted electrodes have been more vastly used for their low cost, high temporal resolution, convenience and non-invasiveness and will be the focus of this work.

Pattern recognition and classification algorithms can then convert the acquired signals into the required commands to translate users's intent. For this purpose, a great variety of features has been tested in literature according to the type of BCI system in analysis: amplitude values, power spectral densities, band power values, autoregressive and adaptive autoregressive parameters, time-frequency features and inverse model-based features are only some examples. In general, when selecting the best features for accurate and stable classification, some critical properties must be considered [16]:

- noise and outliers, caused by the poor signal-to-noise ratio (SNR) of the EEG signal;
- high dimensionality of the feature vectors (features are extracted from several channels and several time segments);
- time information, as brain patterns are strongly time-variable;
- non-stationarity, with ample variations across time and sessions;
- small training sets, as the training is time consuming and demanding for the subjects.

However, feature extraction is only the first step in the software design of a BCI: numerous classifiers have been tested, taking into account accuracy and robustness, as they must be reliable both intersession and interpatient, although not lacking a certain degree of personalization. The main issues that must be addressed in the classifier selection are the curse of dimensionality and the bias-variance trade off. The curse of dimensionality describes the exponential increase in the amount of data necessary for classification when dimensionality of the feature vectors increases: experience has demonstrated that five to ten times the feature vectors dimension is the minimum number of training samples that should be used to avoid poor performances in the classifiers. However, in BCI, dimensionality is often high and the training set small, therefore this is a major concern and results should be evaluated on a case by case basis. Conversely, the bias-variance trade off refers to the three sources of classification errors (noise, bias and variance) and how the bias and variance have a natural trade off, so that the classifier either has a strong divergence between the estimated mapping and the best mapping (bias) or a strong sensitivity to the training set used (variance): stable classifiers usually have high bias and low variance, which can solve the EEG inter-session variability problem of BCI systems.

The classifiers most commonly used in BCI applications are divided into five categories [16]:

- Linear classifiers: the most commonly used, they use linear functions to separate classes. They are well suited to real-time applications and are very stable. In particular, they can be further categorized into Linear Discriminant Analysis (i.e. LDA, which uses hyperplanes that maximizes the distance between the two cluster means) and Support Vector Machines (i.e. SVM, which use the same principle as LDA but applied to the training points closest to the hyperplane). LDA has good performances, but cannot handle outliers and complex non-stationary signals, therefore SVMs are considered an improvement on this model and are the gold standard in many BCI applications. They have good generalization properties, are insensitive to overtraining and to the curse of dimensionality.
- Neural Networks: the second most used type of classifier in BCI applications, they enable nonlinear decision boundaries. In particular, multilayer perceptrons are widely adopted in literature because of their flexibility: they can approximate any continuous function and classify any number of classes. However, they are sensitive to overtraining and careful architecture selection and regularization is required.
- Non-linear Bayesian classifiers (specifically Bayes quadratic and Hidden Markov Model): these classifiers produce nonlinear decision boundaries, are probabilistic in nature and can successfully reject uncertain samples. Nonetheless, these classifiers are less commonly used in BCI applications.

- Nearest Neighbor classifiers: they assign the test feature vector to the dominant class among the nearest neighbors (in terms of training classes). They are not very popular for BCI applications because of their sensibility to the curse of dimensionality and the characteristics of the class clusters.
- Combinations of classifiers, that can be aggregated in different ways and using different types of classifiers.

2.1 BCI classification

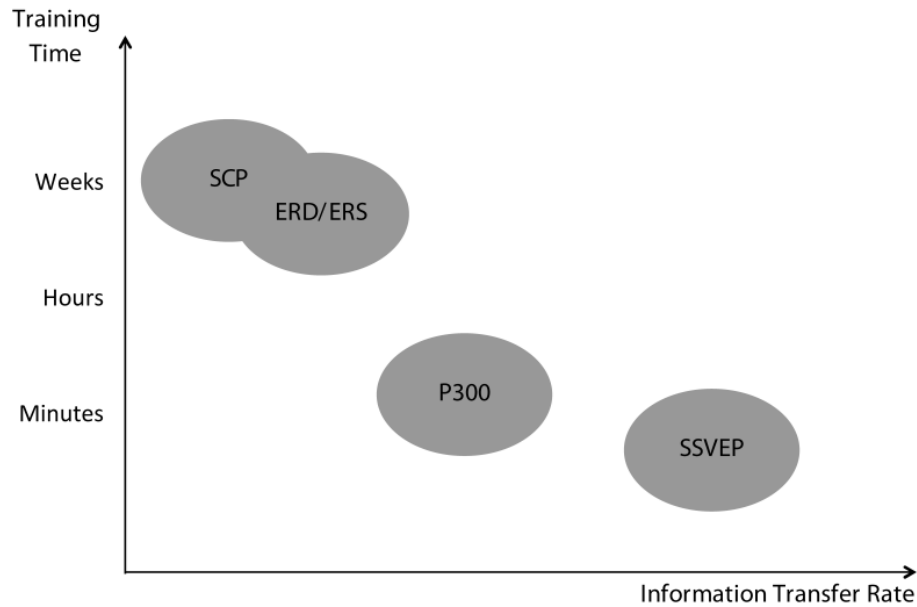


Figure 2.2: ITR and required training time for different BCI systems. Note that SSVEP-based BCIs have higher ITR with little to no training. (Image taken from [2])

Most BCIs currently being studied are based on brain rhythms modulation, in order to allow the patient to exert his or her own volition. Brain rhythms can be directly or indirectly controlled by the users, in a way that does not depend on the peripheral nerves and muscles (which might be compromised in the patient).

In this context, BCIs are categorized into several types according to the EEG pattern used for classification:

- Event-related desynchronization/synchronization (ERD/ERS or motor imagery), which requires imagining movement in order to cause a desynchronization in the motor cortex activity, i.e. exploit the sensorimotor rhythms [17];
- Steady state visual evoked potentials (SSVEP), which are the subject of the present study [18];
- P300 component of event related potentials (ERPs), elicited by the appearance of a low-probability target stimulus, mixed with high-probability non-target ones;
- Slow cortical potentials (SCPs), which is now seldom used because it requires extensive training, but has lower accuracy and robustness than other approaches [19].

Each of these paradigms has different advantages and limitations, as underlined by Figure 2.2. In particular, it is interesting to note that SSVEP-based BCIs have high classification accuracy, high information transfer rate and need little training. On the other hand, they can be tiring for the users because of intrinsic paradigm characteristics (continuous flickering matrix) and require very precise timing of the frames in the proposed scenario of blinking stimuli.

A further classification of BCI paradigms divides them in two classes (as show in Figure 2.3):

- Induced, stimulus independent, e.g. motor imagery and neuro-feedback
- Evoked, stimulus dependent, e.g. SSVEP- and P300-based BCI

Although an interface is always necessary to provide a feedback for the BCI and therefore help the patient better control the system, the induced paradigms have the advantage of not requiring external stimuli. In fact, to control some types of BCI only imagining the movement (in the case of motor imagery) is necessary: this is an extremely natural task for the subject, especially to command a robotic arm or a wheelchair. In some paradigms, control is exerted through mental activity, after the users have learned to modulate their brain rhythms thanks to a feedback and reward system: this eliminates any imprecision or inconvenience caused by a stimulation screen. On the other hand, these systems require a pretty extensive training and may incur in bigger “BCI illiteracy” problems (see the section “BCI user demographics”). In this context, evoked paradigms may be more suitable for a wider range of subjects.

A more in depth comparison between two of the most common evoked paradigms (P300- and SSVEP-based) will be provided in the next section.

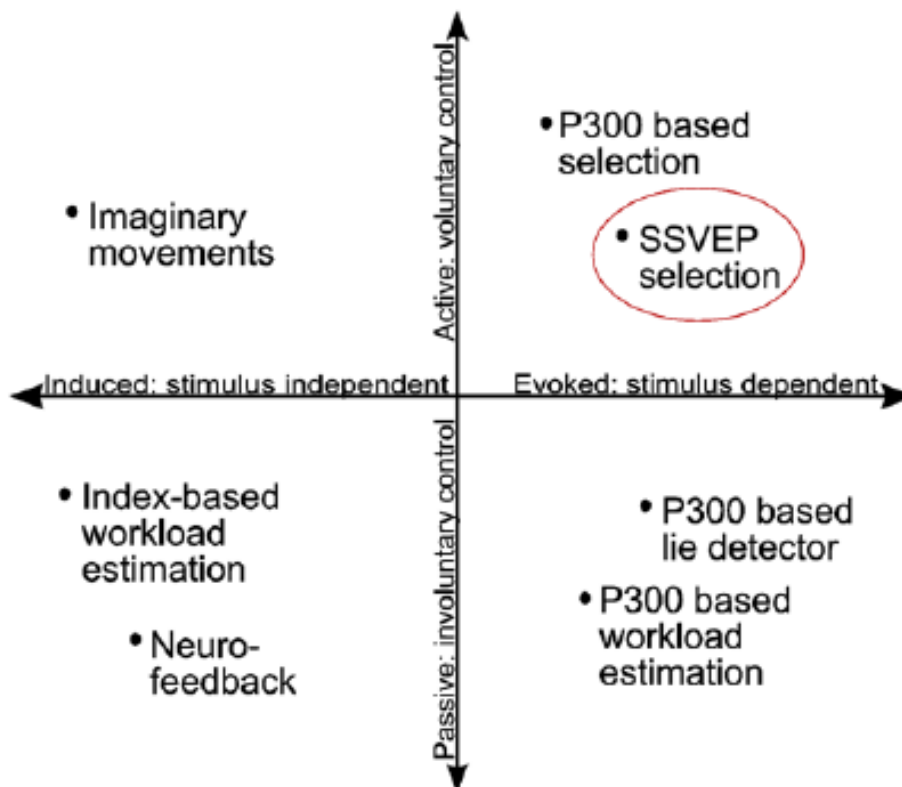


Figure 2.3: Evoked vs induced and active vs passive BCI classification (image from [3])

2.2 Comparison of SSVEP- and P300-based BCI performances

A P300 is a late appearing component of an auditory, visual or somatosensory ERP, with a latency of about 300 ms. It is elicited by rare or significant stimuli, administered in an environment where frequent or routine stimuli are provided. Its amplitude is strongly related to the unpredictability of the stimulus, therefore the surprise factor is an important part in the BCI paradigm. The subject is asked, in a speller application, to focus on a symbol in an agreed list (for example, a letter in the alphabet) and when the symbol is either pronounced by a recorded voice played in an earplug or lights up on a monitor where all the symbols are present, the users will react to the symbol being selected and the P300 response will be recorded. As such, a perfect temporal agreement between the symbol selected by the paradigm and the EEG response recorded is necessary [1].

On the other hand, a SSVEP is defined as a periodic evoked potential caused by the subject gaze being fixed on a flickering target (at constant frequency in the 6-75 Hz range), as seen in Figure 2.4. These responses are composed by a number of discrete frequency components, in particular one fundamental frequency corresponding to that of the visual stimulus and its harmonics. They can be detected non-invasively using an EEG with surface electrodes (e.g., scalp electrodes mounted in an elastic cap), located in the occipito-parietal region, as demonstrated in previous studies [20, 21]. This is due to the fact that the parietal and occipital electrodes mainly record the activity of the primary visual cortex. SSVEPs have the desirable characteristic of being elicited only by flickers inside the visual field of attention, in particular those stimulating the foveal region of the retina, making flickers at other frequencies surrounding the desired one irrelevant [1].

P300 potentials and VEPs allow the design of BCIs with a large amount of commands and high communication speed. Typically, a P300- or VEP-based BCI is capable of encoding more than thirty different symbols (e.g. alphabetic characters or commands). For example, the P300 speller proposed by Farwell and Donchin[22] in 1988 was a 6x6 spelling matrix and reached a ITR of 5 characters/min. In the last two decades improvements on the paradigm [23] and recording [24] allowed higher ITR. However, the speed of a P300-based speller is intrinsically limited by the duration of a symbol intensification and the time delay between two consecutive characters.

Generally speaking, VEP-based spellers are not limited by the same constraints and show a higher communication speed than P300 spellers, up to an average ITR of 20÷30 bits/min in a 26-target system [25]. Furthermore, the SSVEP is a natural response of the central nervous system to visual flickers and therefore a BCI system based on it requires little training for the subject to start using it. However, the accuracy, user variation and

ease of use of the SSVEP-based BCIs found in literature can be improved, if compared to other types of BCI [21].

Specifically, their performance are limited by [6]:

- stimulus presentation: the flickers must be stable and accurate (using embedded hardware or clinical software, as described in chapter 3);
- multiple target coding vs limited number of encoding frequencies: only stimuli in a specific frequency range elicit a SSVEP response. Additionally, the harmonic frequencies of another stimulus cannot be used for the flickers and the number of frequencies is limited by the refresh rate of the monitor used for the stimulus presentation (these problems can be solved respectively by accurate paradigm design and by an approximated frequency approach);
- target identification: efficient identification is required (a multichannel approach and spatial filters such as Canonical Correlation Analysis can improve the classifier accuracy).

High ITRs were also reported in BCIs using other types of VEP signals (refer to [26]).

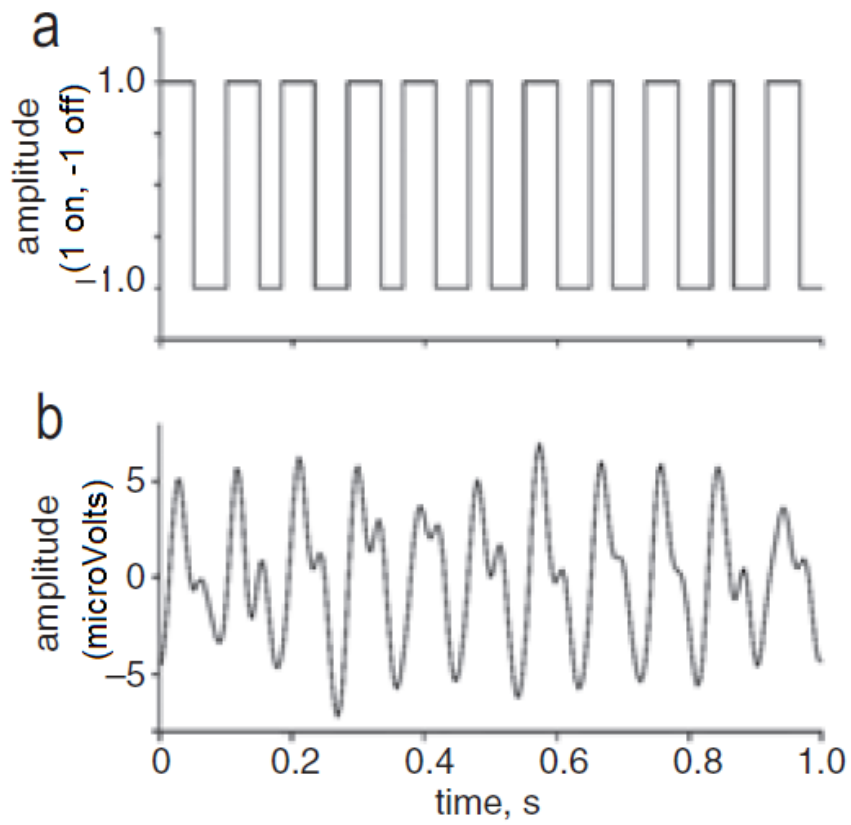


Figure 2.4: Flickering stimulation and consequent SSVEP response. (Image from [4])

2.3 SSVEP-based BCI: State of the Art

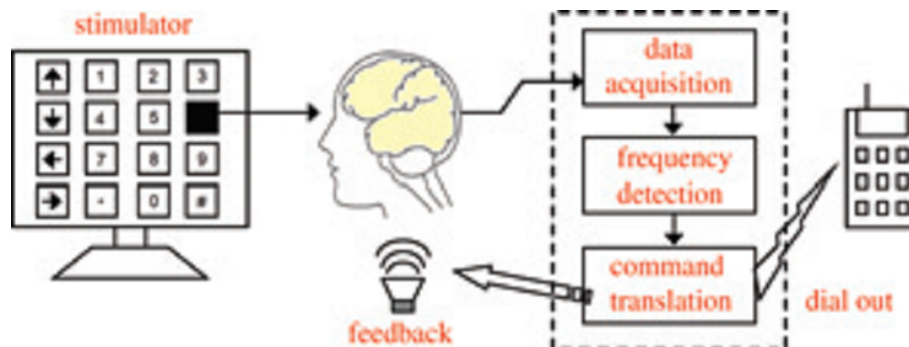


Figure 2.5: Typical setup for a SSVEP-based BCI (Image taken from [4]). Note the stimulation screen, data acquisition, signal processing and biofeedback blocks.

The mechanism behind SSVEPs is not yet well-understood, but recently SSVEP-based BCIs have been increasingly employed in research [27, 28], as they have been demonstrated to be useful in different applications, especially the ones (like communication) that require a large number of commands and self-paced performances.

A criticism that is often moved against SSVEP BCIs is that it requires gaze shifting, therefore its application for users with certain types of locked-in syndrome is not possible and, furthermore, an eye-tracking device would be an easier solution to the problem. However, the literature on visual attention [29] has established that humans can shift their attention without shifting their gaze and other experiments have established that SSVEP responses can be elicited simply through attention shifting, although gaze shifting does improve the performance. It is therefore necessary that a paradigm is developed taking into account these considerations, so that it can be entirely inside a person's field of view without necessarily requiring eye movements that might be compromised.

A typical SSVEP-based BCI system uses a grid of light-emitting diodes (LEDs) that flicker at different frequencies, each representing a different symbol or command (a typical implementation of such a system can be seen in Figure 2.5). Computer monitors have also been used, especially CRT ones because of their higher accuracy in vertical refresh rate (due to an input lag of less than $1 \mu\text{s}$ [30]) and lower ghosting than LCD: these characteristics highly impact both the quality of the recordings and the ease of use for the test subjects. SSVEP responses can be measured within very narrow frequency bands around the visual stimulation frequency, with the minimum frequency variation that elicits different responses in the visual cortex being 0.25 Hz. Therefore, several stimuli can be implemented using a relatively small band of flickering frequencies, usually in one of the three ranges with highest SNR [2]: low frequencies ($6 \div 12$ Hz), medium frequencies ($13 \div 25$ Hz) and high frequencies ($40 \div 60$ Hz).

After acquisition, the raw EEG data is processed to extract the frequency components of interest: for this purpose the most widespread technique is based on power spectral density (PSD) on a sliding time window, as the maximum spectral peak should be centered around the visual stimulus frequency. The PSD can be estimated using the periodogram, which has a low computational cost but is sensible to noise; for this reason, it is important that the data analyzed are recorded from multiple channels [20]. Other methods attempt to improve on robustness upon the PSD estimate, like autoregressive spectral analysis and the frequency stability coefficient (SC) [31], which has been shown to have higher accuracy than power spectrum for short data windows, but at the cost of longer training to build the SC model.

Other techniques like template matching and recursive outlier rejection have also been studied in relation to the feasibility of SSVEP BCI systems. Furthermore, spatial filters have been applied to the data to improve their quality: these filters combine the information from independent electrodes (or channels) to improve the spatial resolution and enhance the desired frequency components. In fact, raw EEG scalp potentials are known to have poor spatial resolution due to volume conduction of the cranium and their SNR might be low. In a simulation study [32], it was proved that only half of the signal recorded by each scalp electrode came from the neuronal population within a 3 cm radius. This is an important issue if the signal of interest has small amplitude, especially when noise sources produce strong disturbances in the same frequency range, e.g. the α rhythm in the occipital region or movement and muscle artifacts. In single-trial analysis (like BCI), it is fundamental to remove these components as much as possible while improving the strength of the desired signals, either by training the subjects or by calibrating the system to the specifics of single users. Among spatial filters, Common Spatial Patterns (CSPs, which separate the multichannel signal into additive subcomponents with maximum differences in variance between two windows [33]) and Canonical Correlation Analysis (CCA, which finds linear combinations of two matrices so that they have maximum correlation with each other) have proved effective. In regard to SSVEP-based BCIs, studies [34] show that the best results are obtained with CCA, which is also an efficient method for an online classification, as the required window lengths are shorter than those necessary for power spectrum estimation. Most CCA methods consist in a correlation maximization between the multichannel EEG signals and reference sinusoidal signals at each of the used stimulus frequencies (and their harmonics) [35]. The frequency corresponding to maximal correlation is then taken as the stimulus frequency: this method has been proved to have significantly better recognition performance than that of PSDA, computed using a single or bipolar channel [21]. Although the CCA approach for SSVEP-based BCIs has been validated by several studies [20], a potential problem arising is that all parameters for recognition come from the test data, because the reference signals of sine-cosine waves do not include features from training data [35].

Therefore, the CCA method can be improved upon using the approach from Nakanishi et al. which exploits cross-correlation with both reference signals and training sets. This new technique will be described in detail in the "Approach" section of chapter 3.

2.4 BCI user demographics

Several studies [29, 36] have analyzed in depth the variability across subjects in efficacy of use of SSVEP BCIs and how the success or insuccess were related to factors like age and gender of the users and their lifestyles. In fact, intersubject variability and the necessity of parameters customization are some of the most consistent observations in BCI literature and also considered major factors in BCI illiteracy, i.e. the percentage of users who are unable to obtain sufficient control of the system. As such, it is important to analyze the demographics of BCI users and which factors contribute to the success or insuccess of a specific approach. Research on demographics of BCI users has been proved to be an effective tool to study and possibly improve the effectiveness of BCI systems.

The empirical conclusions found in literature do not have statistical significance, but seem to point toward higher performances (in terms of ITR) in female subjects and people who regularly play videogames. Furthermore, elder subjects (who have also been demonstrated to have lower SSVEP response) report being more annoyed by the flickering lights. Interestingly enough, there seems to be no difference in performance between subjects who gaze at the target and those who focus on it. Phrase length and free spelling vs copy spelling seemed to have no significant effect on performance either. In this context, some studies [29, 37] have found a 100% success rate (as in, effective communication with an accuracy equal to or higher than 60%), across all healthy subjects: however, this result cannot be generalized to prove that SSVEP-based BCIs are a universal solution, as the problem of BCI illiteracy has been encountered in other studies [36, 38].

Regarding locked-in patients, the relationship between illiteracy and the severity of motor impairment has been analyzed in depth. Using a threshold of 70% accuracy, it has been proved that no correlation subsists between severity of impairment and illiteracy, except in completely locked-in patients because of their inability to shift gaze [39].

It can be concluded that, before experimental trial, it is difficult to predict whether or not a subject will succeed in using a certain BCI approach, which is why it is important to keep in mind that other options (i.e., different BCI approaches, paradigms, signal processing algorithms and parameters) can be explored.

Efforts to improve the universality of specific BCIs have only been partially successful: extensive training of the subject and/or classifier, alternate displays or instructions, improved signal processing efforts and error correction have all yielded better results for specific previously illiterate subjects, some people remain unable to use a particular BCI system [38]. As such, it can be said that there is no universal BCI: a relatively small percentage of test subjects will probably never obtain satisfying results with a given approach. This is due to the fact that, while the same cortical processing systems can be universally found in roughly the same areas with respect to the standard electrode

positioning, there are individual variations in brain structure. It is theorized [36] that, in some people, the cortical area responsible for a specific brain rhythm might not produce electrical activity detectable on the scalp. Even though the neural populations are presumably healthy and active, they may be located in a sulcus, too deep for EEG electrodes or too close to another louder group of neurons.

Chapter 3

Material and Methods

The stimulation interface was created using Presentation® software (Version 18.3, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com), data acquisition and channel selection were implemented using the Matlab toolbox EEGLAB 13.5.4b (Schwartz Center for Computational Neuroscience, UC San Diego). The training and classification algorithms were implemented in Matlab 2016a (The Mathworks, USA).

3.1 BCI paradigm

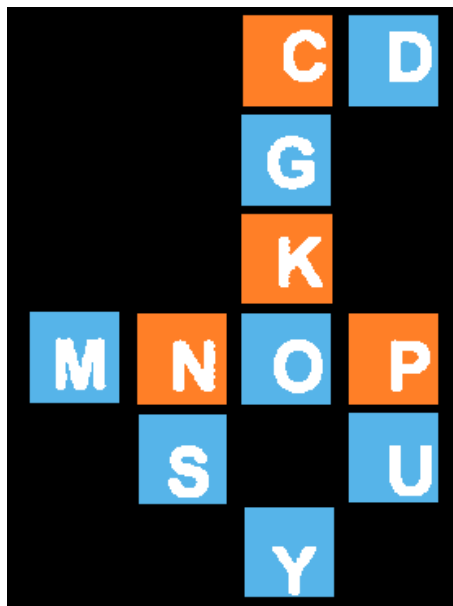


Figure 3.1: A screenshot of the blinking interface. Notice the differently colored squares on a black background.

Paradigm design is a critical step in the development of a BCI system, as it influences its throughput. In particular, in a speller task, the interface must improve accuracy and ease of use, thanks to a user-friendly character layout: this is fundamental for effective communication and to motivate the user [40].

The BCI paradigm (which can be seen in Figure 3.1) was chosen in accordance with the literature studied: a 4x6 typing matrix as per Farwell-Donchin paradigm [22] containing letters from the English alphabet written in white over colored squares, which are placed in a checkboard pattern in orange (RGB code 230, 130, 20) or sky blue (RGB code 86, 180, 233) on a black background. Each square flickers at a fixed frequency, so that the letter itself and its background keep alternating between black and the chosen color. On the other hand, the black borders surrounding the square, between one letter and the neighboring ones, do not change color. The frequency of each letter (which ranges from 8 to 21 Hz, as described in more depth later in the chapter) was chosen so that stimuli close to each other have very different frequencies. All these steps were taken to prevent the user's attention from drifting to other symbols as much as possible, which would attenuate the response in the recordings: this was possible thanks to surveys done on potential test subjects, asking them for feedback on different aspects of the interface, especially what they felt made it easier to concentrate. It is in fact important that the interface is as comfortable as possible for the users and that it facilitates focusing their gaze: attention, motivation and user-friendliness are important factors to obtain high quality signals. Furthermore, visual fatigue decreases SNR, therefore it must be delayed as much as possible. For this purpose, the feedback gathered was used to create an interface that satisfied the highest number of subjects, in particular to identify colors that were easy to discriminate and not tiring to the eye. Specifically, white on black (very common in this type of interface) was avoided because it caused several subjects to report headaches or watering eyes, caused by the high contrast of the blinking stimuli at frequencies that are often found annoying by test subjects [5]. The colors chosen have a good contrast on the black background and similar relative luminance (respectively 63.7 and 64.25), but very different RGB composition; furthermore they are also easy to discriminate for people with all forms of deficient color discrimination, although the totality of the subjects tested did not declare any previous diagnosis of color blindness.

The flashing frequencies were selected considering the range that elicits a SSVEP response and in particular the subranges that have the highest SNR in the EEG recordings of the parieto-occipital region of the cranium, as they have been identified by previous studies [5]. Of the three ranges with highest SNR (reported in Figure 3.2), the one between 5 and 25 Hz was chosen, as there is less background EEG activities at higher frequencies but interferences from the myoelectric signal components become more important and the SSVEP response has smaller amplitude. Especially with recordings from electrodes in occipital positions, muscular activity from the neck and back cannot

be neglected in experiments with healthy subjects, while it becomes less relevant for patients with severe pathologies that impair the neural motor pathways or the muscular system [41].

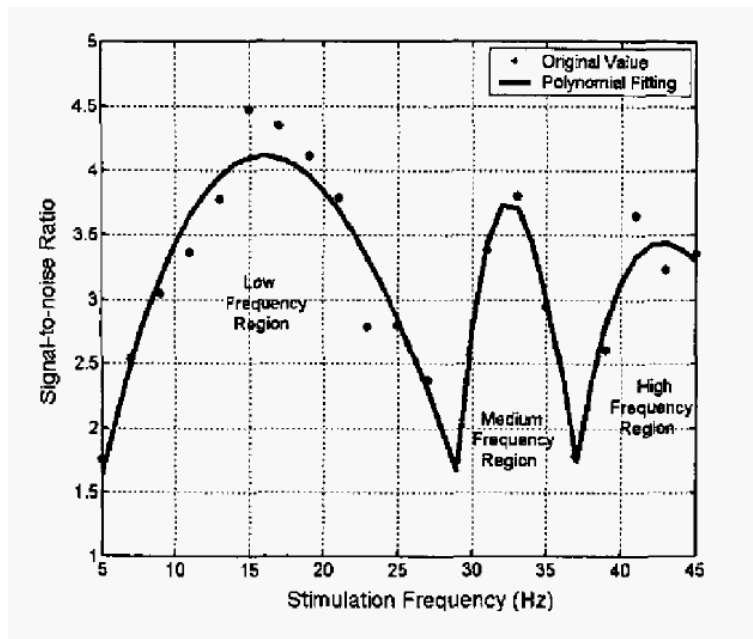


Figure 3.2: SNR with respect to stimulation frequency: note the three subsystems with peak SNR. (Image from [5])

The presence of an alpha rhythm was also considered, as its range partially overlaps with the most used frequencies in this type of systems and is recorded in the same area of the scalp as the SSVEP response. The alpha rhythm is prevalent when the subject is relaxed and with closed eyes, however experimental trials during the project demonstrated its presence for certain subjects during task execution (which will be discussed more in depth in the following sections). Since the range for the alpha rhythm is usually between 9 and 12 Hz, flickers in this range were avoided, so that the SSVEP would not be contaminated by the alpha rhythm. In fact, simulations proved that signals at less than 1 Hz of the main alpha component induced classification errors, especially when the the alpha rhythm had Fourier transform with amplitude close to that of the SSVEP component (see chapter 5).

Lastly, the relationship between the stimulation frequency and its harmonics or sub-harmonics was taken into consideration: the nonlinearity of the central nervous system causes important harmonic components in the SSVEP response. It is therefore impossible to use a certain frequency and its harmonics as distinct stimuli as they elicit the same SSVEP response. Some studies [42] have found that for certain subjects the second

harmonic component has even higher amplitude than the fundamental frequency: this must be considered in the training of the classifier (for example, responses at both 8 and 16 Hz should be considered as the same class by the trained algorithm). As such, a series of square wave impulses with frequencies of 8, 9, 13, 15, 17, 19 or 21 Hz (which are within the region with highest SNR for the SSVEP response [4]) were used. A series of trials used different combinations of the selected frequencies.

The interface was created in order to maximize the dimension of each of the 24 squares and the distance between them: however, 24 stimuli (the maximum number deemed acceptable, considering the screen in use for the experiment) are not enough to cover the entire English alphabet. Since the paradigm was implemented for experimental purposes only, i.e. to investigate the ability of the classifier to discriminate the chosen frequencies that were within the desired range, the letters q and x were not present in the interface. For this reason, the interface should be further modified in the future to include more stimuli (see chapter 6), possibly optimized for wider screens. However, in a final implementation of the interface, it should be taken into consideration that a smaller alphabet also helps with the classification accuracy: the so-called “Italians Do It Better” Effect. In fact, the English language uses 26 characters, while Italian uses only 21 and, in a virtual keyboard task, selecting a character for an Italian-speaking user should be easier with respect to an English-speaking one: respectively, to select each symbol, 4.392 bits are required, instead of 4.700 bits [43]. As such, the interface could avoid implementing certain letters and agree with the users on some conventions, for example substituting diphthong “qu” is substituted by “kw” and the letter x by “ks” during a free-run spelling task. Furthermore, a spell checker algorithm could, after classification, automatically recognize and operate the inverse substitution (see chapter 6).

It must be taken into account that SSVEP-based paradigms have an important limitation, as discussed in chapter 2: flashing frequencies are limited by the technical characteristics monitor. Since the refresh rate of the CRT monitor used was of 60 Hz, certain frequencies should not be possible to reproduce, as the color reversal would have to occur after a non-integer number of frames. Theoretically, only frequencies submultiple of 60 could be fully implemented (e.g., 6, 10, 15 Hz). Therefore, to have a wider range of possible encoding frequencies (i.e. more stimuli), a different approach could be used: a method has been proposed and validated in recent years [4] to generate flickers at any frequency up to half the monitor refresh rate, by approximating the frequency itself with a variable number of frames in a stimulation cycle [6]. The method essentially creates a square wave with duty cycle of 50% and variable frequency: alternating every few frames between two different frequencies, the resulting impression to the viewer (and therefore the elicited evoked potential) is of just one oscillation frequency, with value between the two real ones. This depends highly on the human visual system, which is not equipped for such fine discrimination of frequency variation (in the order of a few dozens of ms,

in the typical situation of a monitor refresh rate of 60 Hz). An example of the resulting stimulus is seen in Figure 3.3: the cycles in a second are the same as a non-approximated approach, but a change in a non integer number of frames is not required.

Therefore after defining the borders of every flickering area, a series of one-second long square waves were created in Matlab according to the formula:

$$s(f, \phi, i) = \text{square}[2\pi f(i/\text{RefreshRate})] \quad (3.1)$$

Where s is the square function, f the chosen frequency and i the discrete time. The duty cycle is fixed at 50%. When the square wave has a rising or falling edge, the corresponding area on the typing matrix should switch (from black to color, and vice versa).

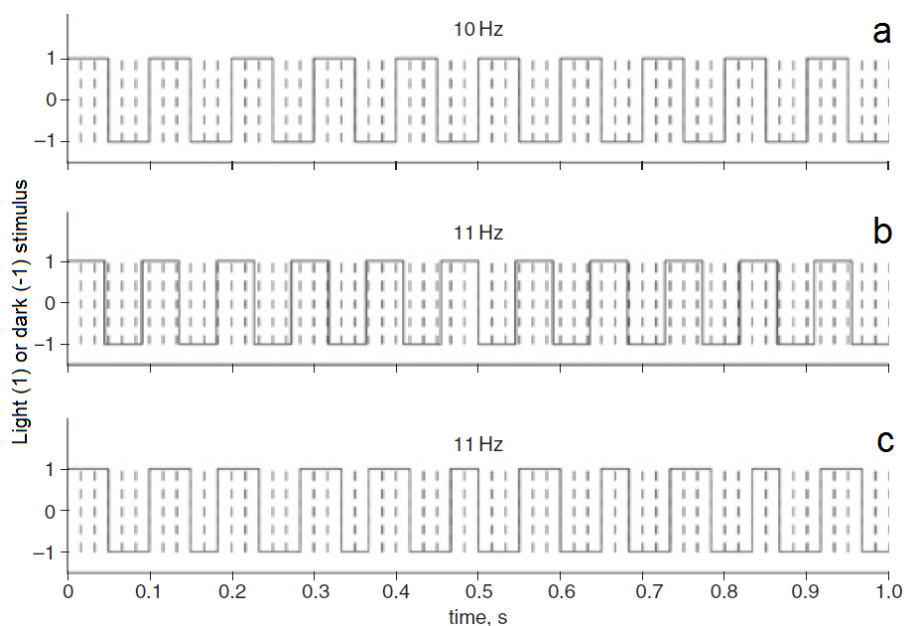


Figure 3.3: The approximation approach: considering a refresh rate of 60 Hz, a 10 Hz frequency (picture a) can be easily implemented by switching every three frames. On the other hand an 11 Hz frequency would require switching every 2.73 frames, which is obviously impossible. Therefore, we alternate between switching every 2 or 3 frames, by using Equation (3.1). (Image from [4])

The square waves were created for one second of recording (i.e. sixty frames), with the previously mentioned approximated approach, were then imported in Presentation where the scenario was programmed in the proprietary Java-based object-oriented language. Twenty-four different signals were created in a way that each square flickers independently from the others; after one second of recording the square wave is looped back to the beginning, in a way that allows a free run of the presented scenario for as long as necessary. Presentation as a software is used by the medical community for BCI applications and for evoked potentials purposes: it has the ability to gain control over the computer resources, using a unique micro-threaded design, and, especially for newer higher-end computers with multiple cores and high clock frequency, can present accurate stimuli at processor level.

Presentation always synchronizes the appearance of picture stimuli with the vertical scan of the display adapter, i.e. indication from the display adapter device driver that the vertical scan has started, which guarantees that the entire picture stimulus appears on the same vertical scan. However, flat panel displays and projectors can have considerable delays (20 ms or more) before the image appears: the delays are intrinsic in the operation of these devices and cannot be circumvented by software. As such, it is crucial to use a monitor that does not introduce delays in the vertical refresh rate: as CRT monitors do not possess a memory element that has to be discharged and charged at every change in color of the pixel, their refresh is much more stable and several orders of magnitude faster than required. Empirical studies have demonstrated that the Presentation event and the actual start of the scan will be delayed of less than 0.2 milliseconds in CRT monitors [44]. For this reason, Presentation projected the scenario on a 15-inch CRT monitor connected to the computer, placed in front of the subjects at eye level: its refresh rate was checked before and after each test and consistently performed at 59.9 Hz, a deviation too small to cause any problem in the resulting recording, in particular because extremely stable across trials. At the end of the scenario, the software also displays a summary of performances, which details the duration of a single frame, the timing accuracy and what stimulus code was presented at that specific time. A manual overview of several runs of this scenario gave reassuring results: except for the first and last second of the scenario, the stimuli were very stable and accurate, with each frame being refreshed after 16.6 or 16.7 ms with an inaccuracy of 0.1 ms (the exact timing would be, for a refresh rate of 60 Hz, 16.666 ms).

A second paradigm was created using Presentation, where a single bigger orange square blinked at the center of the screen. Each time the operator clicked from a distance with an input device (a standard mouse), the stimulus presentation was paused, only to reprise at a second external input, with a higher blinking frequency. The procedure was repeated for all frequencies used between 8 and 21 Hz, for 20 seconds of recording each time. This paradigm was used both to establish the frequencies with highest SNR and

whether, in this specific setup, the minimum Δf variation in frequency that produces a detectable difference in the spectrum was lower than the variation between the stimuli implemented: although literature on the topic often proposes a Δf of 0.25 Hz, it does not take into account possible imprecisions of real time computing on a general-purpose system. This was also useful to further study the difference in response for different subjects, although SSVEP response is somewhat stereotyped. Finally, the data recorded across three healthy subjects (two female and one male) with this second paradigm were used to create the training set, because these recordings had higher SNR, thanks to the lack of external stimuli which might deviate attention from the target.

Using these paradigms, the test subjects were therefore asked to focus on one square for 20 seconds starting after vocal instructions from the experimenter and avoiding any unnecessary movement, in particular contraction of the facial muscles, which introduces artifacts with amplitude strong enough to completely hide the presence of any evoked potential. After each recording, a short break was taken to avoid eye fatigue.

3.2 Experimental Setup

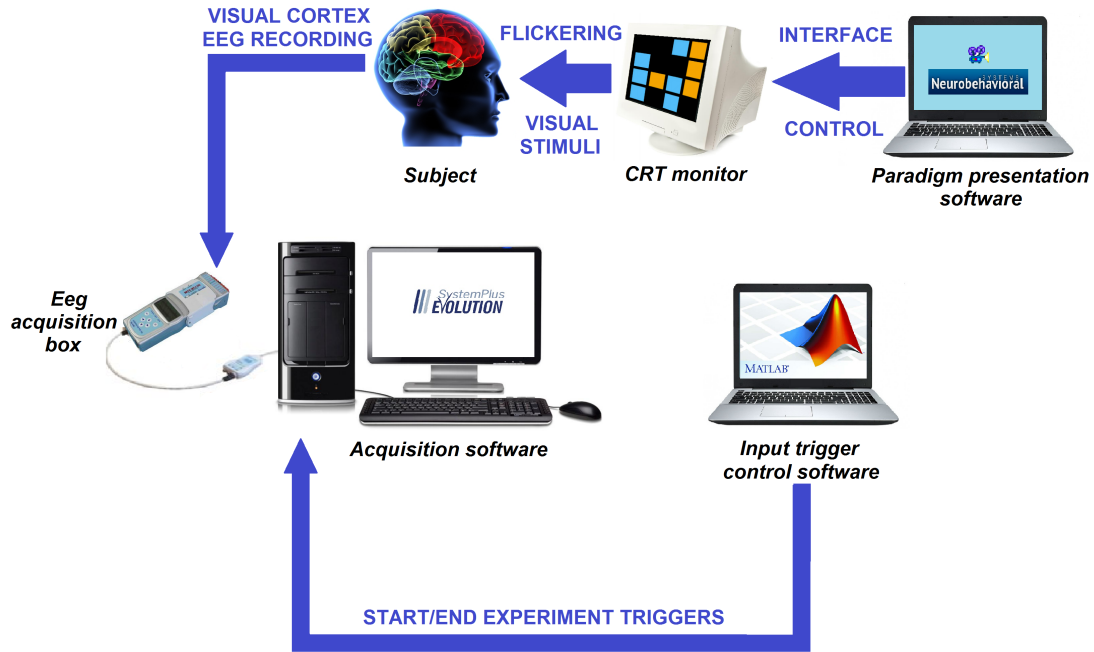


Figure 3.4: Schematics of the experimental setup used during this project.

The recordings were conducted using an EEG cap, acquisition and amplification system and acquisition software from Micromed (Mogliano Veneto, Italy), at Politecnico di Milano. The setup (depicted in Figure 3.4) included:

- A textile customized pre-cabled cap KIT CAP SPEXT61-6: cap and premounted electrodes according to the extended International 10/20 system. The system includes 61 electrodes plus ground and common reference electrodes, 65 labeled cables each 1.5 m long, with touch proof connectors (for the 61 channels, ground, reference and two spare emergency electrodes). The surface EEG is recorded using the mounted gel-based Ag/AgCl electrodes; the electrodes are active, i.e. containing an amplifier inside, so that the gel is injected between the electrode material and the skin to get a low electrode-skin impedance. This allows for a faster montage of the electrode system than with passive electrodes, since it does not require skin abrasion to lower the impedance. In addition, these electrodes are not affected by stray capacitances and cables crossing. One of the main advantages of gel-based electrodes is their robust behavior, but the main disadvantages are the relatively long gel injection procedure and the need to wash the user's hair after the

recording. The setup used for the experiment included the fitted cap, 10 Ag/AgCl electrodes in unipolar configuration (8 recording electrodes: PO3, PO4, PO7, PO8, POz, O1, O2 and Oz, ground fixed on the cap between Fz and Fcz and reference between Cpz and Pz), the chin strap to fix the cap in the calculated position so that each electrode corresponds to the correct area of the cerebral cortex and a thoracic belt to minimize movements that could compromise the recording quality.

- Kit SD KTM 64 EXPRESS: multi-modal biosignal amplifier and digitizer with USB interface, has 72 channels for unipolar and bipolar acquisition of EEG signals, acquisition of EMG and EKG signals. The sampling rate is up to 2048 Hz/Channel; in the present project a sampling frequency of 256 Hz was assumed sufficient for the study purposes, as a 4-second window period was analyzed [6].
- The front-end system was then connected to a computer that acquired and stored the recordings for later use: this computer was separated from the one controlling the CRT monitor, to avoid interference between the two computationally demanding tasks of stimulus presentation and signal acquisition. The latter task was performed using Micromed System Plus Evolution: this acquisition software is fully GUI-based and allows user-defined configuration and setup of the system. It also takes control of the amplification and digitalization of the analog signal. Signals and parameters can be checked in the display mode, stored to disk, and later reviewed in the offline/replay mode. The software allows for serial communication with external triggers, in this case the keyboard of a third computer.

The subjects were asked to sit on a chair at 70 cm from the CRT monitor, their position and the brightness of the monitor adjusted for their comfort so that the display was at eye-level directly in front of them. Afterward, they were strapped to the chair with a thoracic belt around both the chest and the arms, adjusted to minimize movements and relieve part of the load from the postural muscles, allowing the subjects to relax against the back of the chair. This was considered the best way to avoid added muscular noise, both because of unwanted movements and excessive tension caused by long-term sitting. The cap was then mounted on their heads using the Cz position as a reference for correct placement and the electrodes of interest were injected with gel. The impedances and signal quality were manually checked in a closed eye task through the software (if the impedance coupling is satisfying, an alpha component should be clearly visible in the temporal domain). The final setup can be seen in Figures 3.5, 3.6 and 3.7.



Figure 3.5: A view of the experimental setup. In particular, a clear image of the cap and the straps, plus the wired headbox.



Figure 3.6: A view of the experimental setup. Presentation projects the scenario on a CRT screen and the subject is strapped to the chair in front of the monitor.



Figure 3.7: A view of the experimental setup. The whole setup, including the acquisition computer and the computer for trigger generation, can be observed.

3.3 EEG Simulations

EEG simulations were created in Matlab for a first test of the training and classification algorithms: 4 seconds of simulated data with 256 Hz sampling frequency for eight channels were modeled, taking into account the characteristics of real EEG recordings for SSVEP tasks. The sinusoid at the desired frequency and its double (corresponding to the stimulation frequency and its second harmonic) were created for each stimulation frequency. White gaussian noise was then superimposed to them, to model background neural activity. The SNR of the sinusoid with respect to the white noise was imposed to be a random value between 3 and 4 for each trial, as these are values found in literature (see Figure 3.2) for the SNR of an eight-channel recording of SSVEPs in the frequency range of interest, when a recording of a few seconds sampled at 256 Hz is considered. Furthermore, two random sinusoidal distortions and one representing the alpha rhythm (fixed at 10.5 Hz, a reasonable mean value across subjects) were added, although with amplitude smaller than the signal and its second harmonic. This was done to test whether the classification algorithm could deal with spectral distortions, as the real recordings can be expected to be corrupted in certain cases, even when the SSVEP remains detectable. The EEG spectrum has also been proven to have a $1/f$ shape, that highly influences classification over a large frequency band (as will be discussed more in depth in the following sections). However, in the present study the band of interest is relatively narrow and therefore the $1/f$ shape of the spectrum can be neglected. After being generated (examples are shown in Figures 3.8 and 3.9), the simulations are processed and classified using the same techniques as the real EEG data.

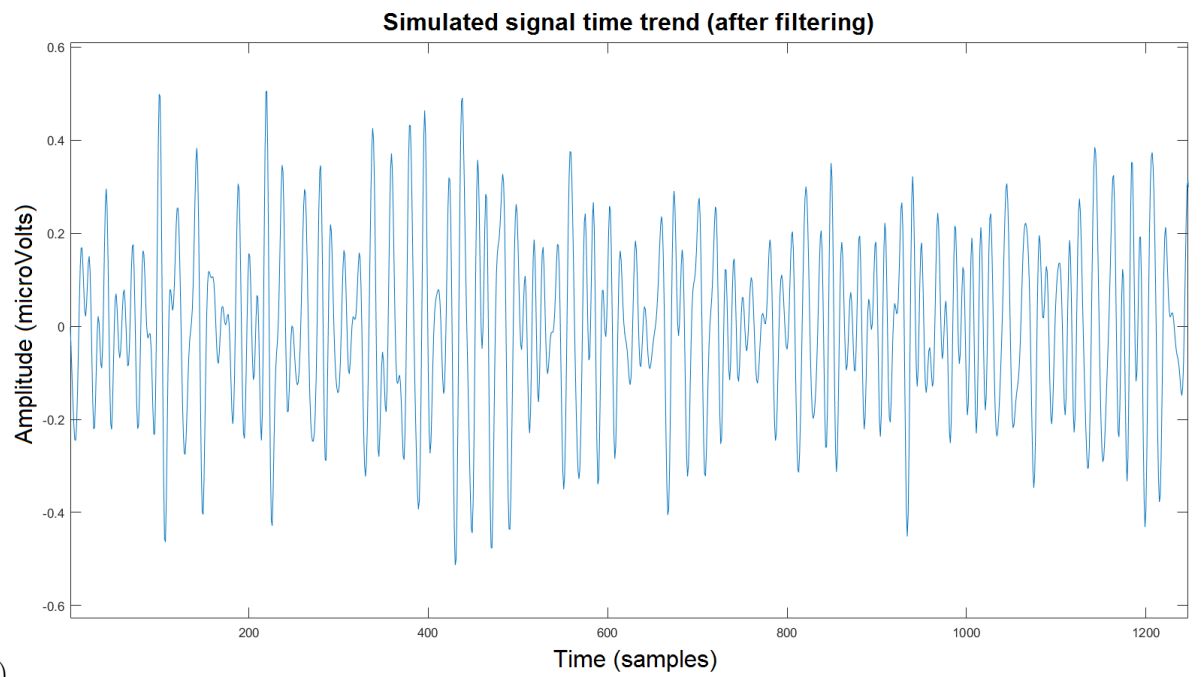
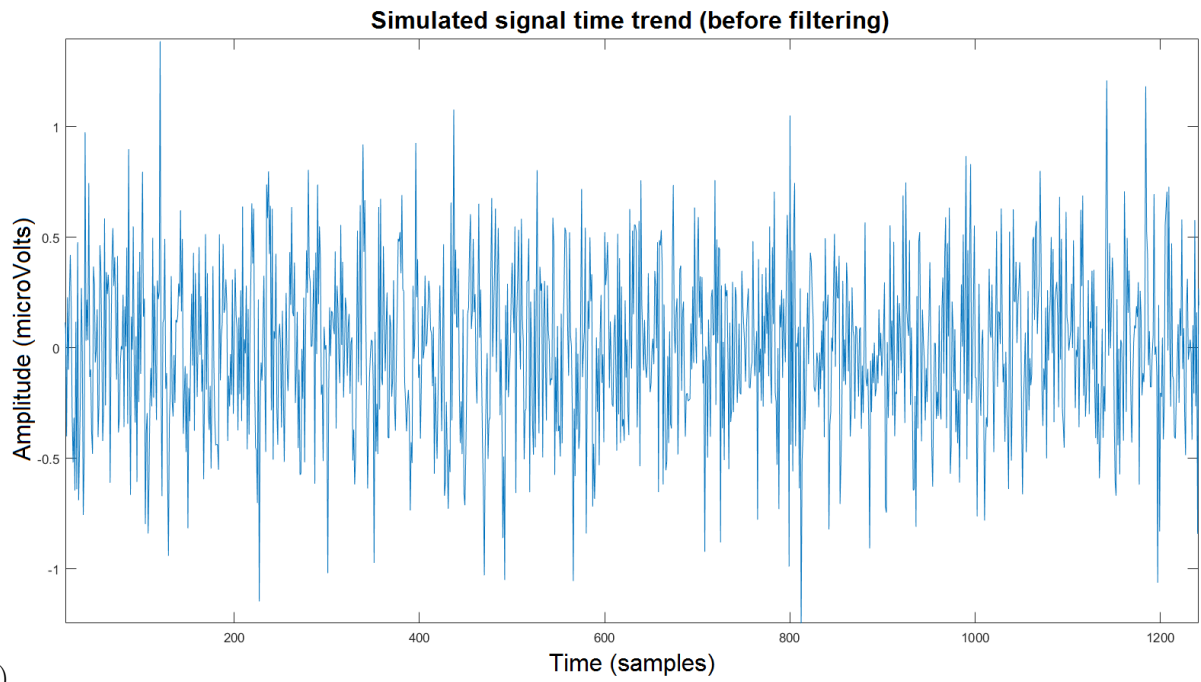


Figure 3.8: Time trend of the simulated signal at 13 Hz before (A) and after (B) filtering.

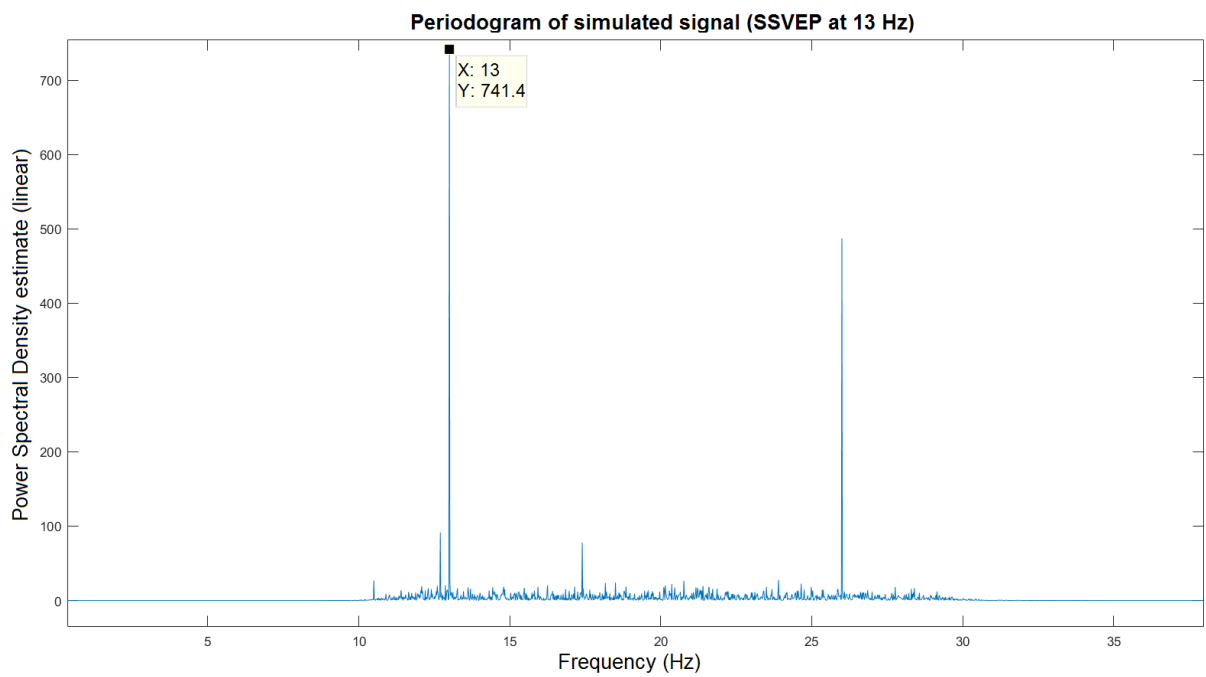


Figure 3.9: Power spectral density estimate of the simulation with signal at 13 Hz.

3.4 Test Subjects Recruitment

Most BCI studies do not employ pre- and post-experiment questionnaires to explore the subjects demographics and their preferences, but it has proven beneficial in some cases [36].

For this reason, when the prospective test subjects were first contacted, they were informed about the procedure details and asked to report any of the following problems (in which case they would be immediately rejected): epilepsy, extreme light sensitivity, skin allergies, history of seizures, predisposition for auras and migraines. This is due to the prolonged exposure to flashing lights at frequencies that could cause epileptic seizures or headaches in sensitive subjects. Afterward, they were asked to sign an informed and free consent form and to fill out a questionnaire right before the start of the experiment. In the questionnaire they were asked about their age, gender and education for demographic analysis purposes; weekly hours of computer work, numbers of hours slept the night before, consumption in the last 24 hours of any substance that might alter their degree of alertness (caffeine, nicotine, psychotropic pharmaceuticals, etc.), to justify the experimental results obtained, as it is well known that the users's ability to focus is essential for the success of the system. Furthermore, questions about their health were asked (as chronic illnesses and mental disorders can influence the result of the system).

After the experiment, the subjects were also asked to fill out another questionnaire, which asked to grade several aspects of the system, using a 1-5 scale system (1 very little, 5 extremely): whether they were under the impression that the system was working correctly (independently on the actual accuracy), if they thought that the system could be of use in clinical applications, their ease at concentrating and shifting gaze on the target, the unpleasantness of the stimuli, the mental effort required to comply to the task and other notes, suggestions or comments to further improve the experimental setup with the users's feedback. These two questionnaires were based on the 2010 SSVEP BCI demographics by Brendan Allison et al [36]. To preserve the privacy of the subjects (considering the sensitivity of part of the information given out), the questionnaires were anonymous, simply associated by an alphanumeric code to the labeled recordings, so that perceived accuracy from the questionnaire and actual accuracy of the system could be correlated with the subject lifestyle and general well-being.

17 subjects were recruited for this experiment, 11 male and 6 female, with mean age 23.2 ± 2.1 years. Out of them, three subjects were called back for a second test, in order to compare the subject's performance across trials.

3.5 Data acquisition and preprocessing

The .trc files, obtained with the Micromed software and containing the recordings, are imported in Matlab and then converted to .mat structures using EEGLAB, an interactive Matlab toolbox designed for processing continuous and event-related EEG, MEG and other electrophysiological data. This allows for the conversion of multiple files and creates a structure containing a variety of information about the recording, among which the raw data from the EEG recording in microVolts (μV), the sample rate of the data (the same for all channels), the trigger times for each channel acquired with the instrumentation and the names of the trigger channels. A channel map was created for the 64 channels of the headset and mapped onto a 4-shell spherical model simulating brain, dura, skull and skin, using the BESA dipolar fitting to localize the recorded data on the extended 10-20 system (Brain Electrical Source Analysis software, from MEGIS Software GmbH, Munich). Subsequently, only parieto-occipital and occipital electrodes are considered, as the other channels had been left open during acquisition and had only acquired electric noise.

The raw EEG data are then bandpass-filtered using a Butterworth filter, with cutoff frequencies of 10.5 and 30 Hz, so that the meaningful frequency range is preserved without attenuation. The filter order highly influences the result: higher order means a sharper frequency cutoff, but the resulting signal might be shifted in time, lower numbers mean a soft frequency cutoff, but the resulting signal is less phase shifted. The filter order was therefore empirically chosen as 7 (the forward and backward filter response can be seen in Figure 3.10) and the recordings were then filtered using the `filtfilt()` function, which performs zero-phase digital filtering, as seen in Figure 3.11.

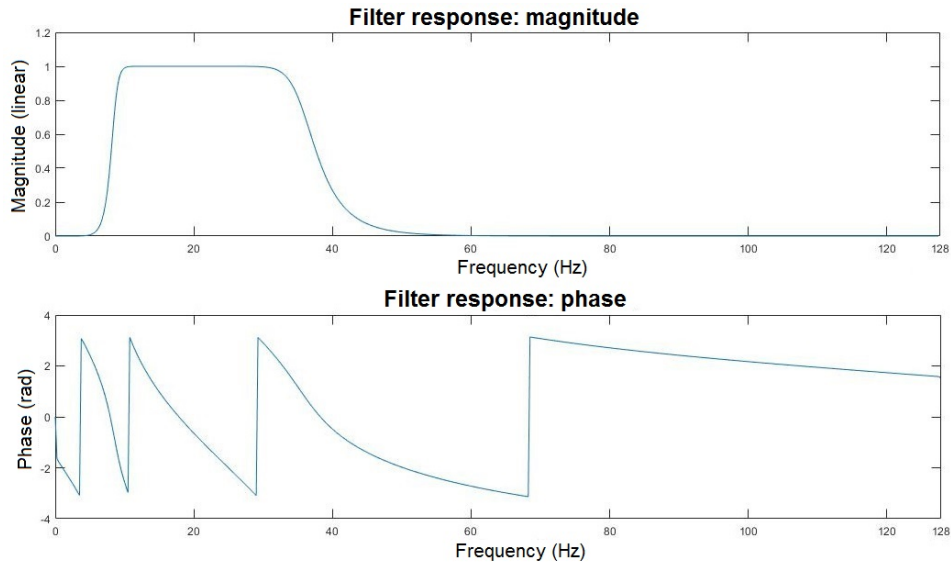


Figure 3.10: Butterworth seventh-order filter: magnitude and phase response. As pictured, the filter has a flat magnitude response in the band of interest, with small transition bands. Furthermore, the phase is more linear than that of other filters of similar nature (e.g. Chebyshev I and II).

First, the 20-second long eyes-closed followed by 20-second long eyes-open recording is studied to extract the alpha rhythm of the subject. Using the eyes-closed recording, the main alpha frequency component is calculated using a 1024-point Welch's spectral estimate. The windowing chosen uses nine 4-second long Hanning windows with 50% overlap (i.e., the spectrum is estimated from the average of nine periodograms calculated on 4-second windows): this has been shown in fact to be the best spectral estimate, with the least leakage both in experimental results and according to literature [45]. Then the highest peak in the 8.5÷12 Hz range is selected as the alpha rhythm frequency. No thresholding method is used, as the amplitude of this rhythm varies highly from subject to subject (usually higher in women) and the threshold should be set ad hoc for each person. Regardless, the value found by the algorithm was manually checked in the power spectral density plotted estimate and found correct in 100% cases. The rhythm is then analyzed subject by subject to understand if the chosen stimulation frequencies are easily discriminated against the alpha frequency (this topic will be discussed more in depth in chapters 5,6).

The processed data are then divided in time sequences of 4 seconds each (which is the duration of a single trial in the training and classification algorithm) and then studied to identify the best electrodes for classification. The channel selection is based on a form factor algorithm created specifically for this project: the SSVEP spectra were ex-

perimentally observed to have very narrow peaks close to the stimulation frequency and amplitude varying greatly across electrodes and some additional peaks of similar width at different frequencies, usually present on all electrodes but with smaller amplitude on some channels. Therefore, the electrodes were chosen for each subject on a training set recorded while the subject was gazing at a blinking stimulus at 8 Hz. It is in fact known from literature that the SSVEP response is more evident on certain subject-dependent areas of the parieto-occipital region and that they are consistent on all frequencies [23]. As such, the algorithm chose the N number of electrodes with highest ratio of power density between the stimulation frequency and the second most prominent noise component. N was varied between 1 and 8, in order to analyze the classifier performances when the number of channels used is increased. In particular, a higher number of electrodes gives more information but is less robust to noisy and drifting electrodes, while a smaller number of electrodes might contain less information or be more sensible to noise: the best trade off was found to be 3 electrodes and this parameter value was used for the entirety of the study (see chapter 5).

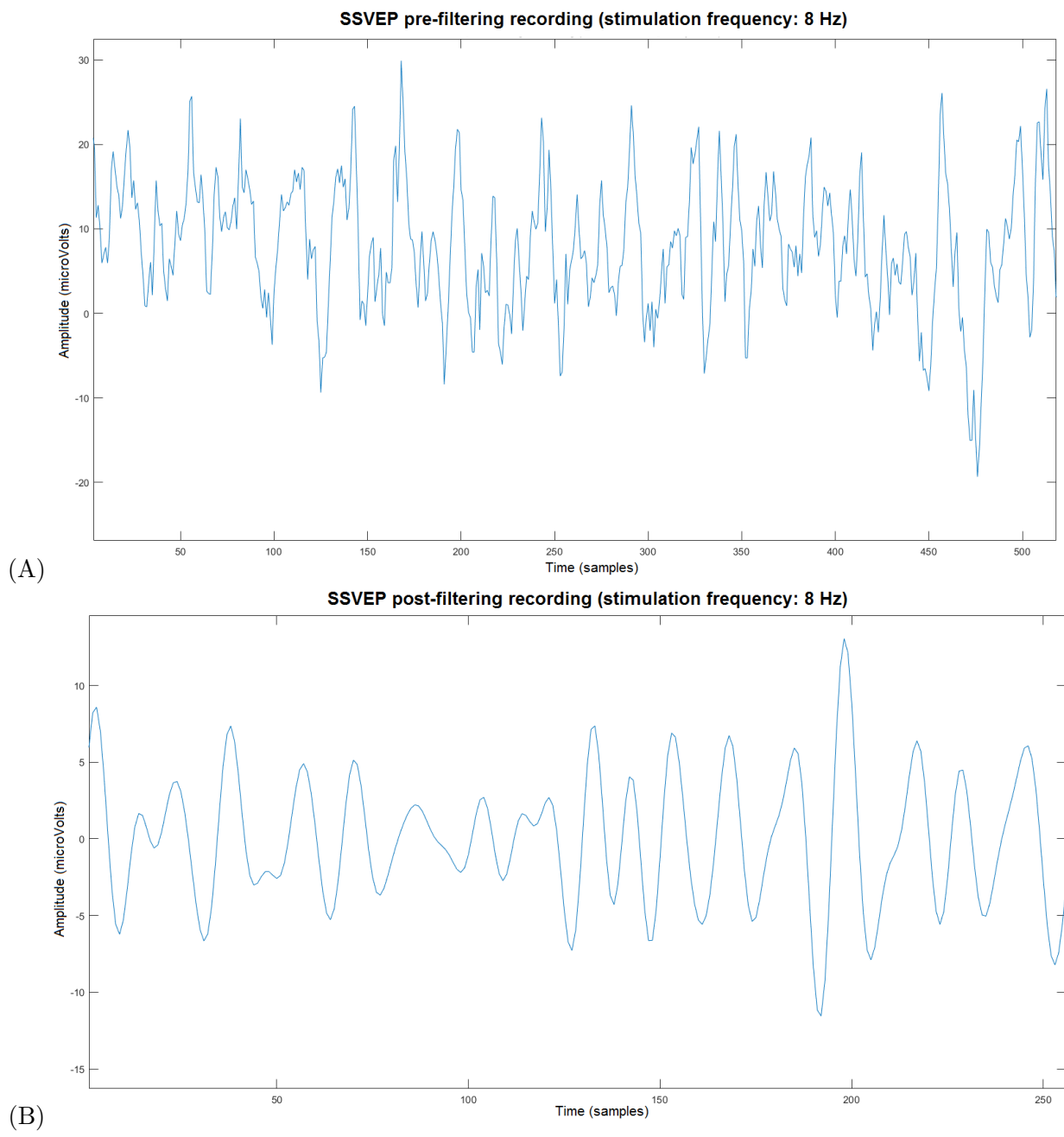


Figure 3.11: Time trend of an EEG recording at 8 Hz before (A) and after (B) filtering. The second harmonics at 16 Hz is strongly noticeable.

3.6 Approach

As stated in chapter 2, a technique that can be used to improve the SNR of SSVEP data is Canonical Correlation Analysis, first introduced by Lin et al in 2006 (more details on CCA and how it is performed in appendix A). CCA is in fact a spatial filter that computes coefficients for each channel of the test recording of unknown frequency and a data set of known SSVEP frequency, so that dimensionality is reduced by combining the information from the channels of each recording according to the computed weight. The correlation coefficient between projections of test set X and training reference signals Y using CCA-based spatial filters can be then used for classification. This allows for a maximization of the Pearson coefficient calculated between the two sets and defined as:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \quad (3.2)$$

Where cov is the covariance between the two variables (i.e. the two sets) and σ is the standard deviation of each variable. As such, the Pearson coefficient is calculated between the fixed test recording and each combination of sinusoidal signals (at a specific stimulation frequency); the frequency corresponding to maximal cross-correlation is then taken as the stimulus frequency. This method has been proved to have significantly better recognition performance than that of power spectral density analysis computed using a single unipolar or bipolar channel when correlation is computed between the test set and reference ideal sinusoids [21]. Although the CCA approach for SSVEP-based BCIs has been validated by several studies [20], a potential problem arises from all parameters for recognition coming from the test data, because the reference signals of sine-cosine waves do not include features from training data [35]. Therefore, the CCA method can be improved using the approach from Nakanishi et al [6] where the CCA coefficients are calculated between combinations of the test set X , training set \hat{X} and reference sine-cosine signals Y , taken in pairs: the average of the Pearson coefficients calculated between combinations of these sets is then maximized to identify the stimulation frequency.

Training reference signals for each symbol in the typing matrix can be obtained by averaging multiple trials in a training set, obtaining \hat{X}_k where k indicates the symbol (with known SSVEP frequency). Afterward, two canonical coefficient vectors are obtained from each CCA performed (between test set X and train set, between test set and sine-cosine reference signals and between train set and sine-cosine reference signals) and are used as a spatial filter to improve the SNR of the data. The correlation vector

ρ is defined as follows:

$$\begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \rho_4 \end{pmatrix} = \begin{pmatrix} \rho(X^T W_X(XY), Y^T W_Y(XY)) \\ \rho(X^T W_X(X\hat{X}), \hat{X}^T W_{\hat{X}}(X\hat{X})) \\ \rho(X^T W_X(XY), \hat{X}^T W_X(XY)) \\ \rho(X^T W_{\hat{X}}(\hat{X}Y), \hat{X}^T W_{\hat{X}}(\hat{X}Y)) \end{pmatrix} \quad (3.3)$$

Where W indicates a canonical coefficient vector. Lastly, a weighted correlation coefficient $\hat{\rho}_k$ is used as the feature for target identification:

$$\max\{\hat{\rho}_k = \sum_{i=1}^4 \text{sign}(\rho_{i,k}) * \rho_{i,k}^2\} \quad (3.4)$$

where $k = 1, 2 \dots, N_{symbols}$ and $\text{sign}()$ is used to retain the information of negative or positive correlation. The frequency of the training reference signal \hat{X}_k for k maximizing $\hat{\rho}_k$ is selected as the one corresponding to the target [46, 6]. The work flow of the CCA filtering and cross-correlation computation can be seen in Figure 3.12.

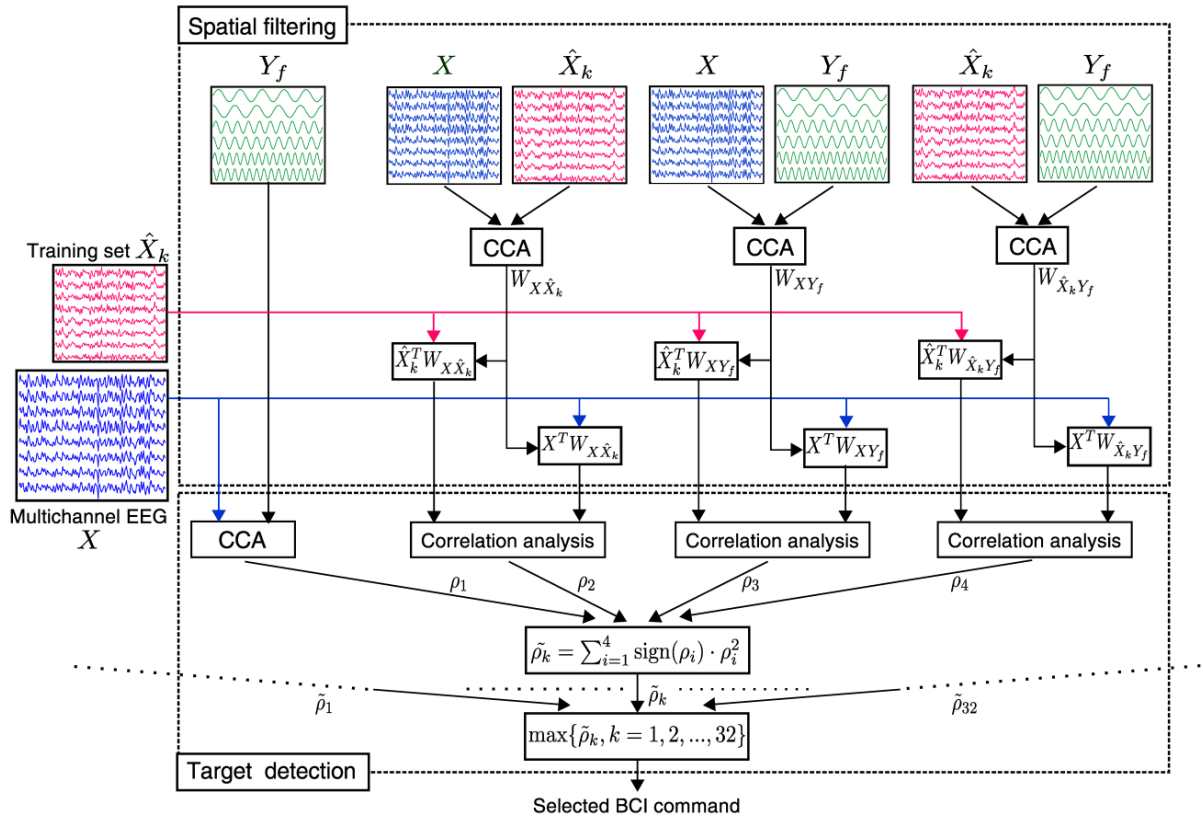


Figure 3.12: Workflow of CCA-based classification, as used in the project. (Image from [6])

However, as seen in the definition of the correlation coefficient, this method is not robust to phase shifts; for example, if the subject is freely gazing at the stimuli and not following a queue (visual or acoustic) to start the task. In fact, by definition of covariance, the sum of the product of samples of the two variables is taken at zero time lag. Two sinusoids of same frequency and no phase shift will have maximal zero-delay cross-correlation (their peaks will coincide and be multiplied by each other, their zero crossings will also coincide and result in less null terms), but two sinusoids at the same frequency with a different phase might cancel each other out. On the other hand, methods based on spectral features extraction are less prone to error in case of phase shift and should be taken into consideration. In particular, if the SSVEP amplitude is much higher than the other recorded EEG components, the power spectral density will have its strongest peak at the stimulation frequency and a simple criterion of thresholding will isolate the correct peak. However, especially at frequency higher than 15 Hz (as seen experimentally in this project), there might be components with amplitude comparable to the evoked potential itself. This could cause equivocation in a threshold-based classifier: as these components have a tendency to appear constantly across trials, spectral similarities could be identified between test and training records.

In particular, the novel approach proposed in this project is based on spectra calculated using autoregressive (AR) models, since they are less prone to spectral leakage than periodograms and the number of samples needed for accurate spectral estimation is lower than that of Welch's method. This estimate would in fact require a sizeable number of windows to average, in turn needing long recordings that would cause a lower Information Transfer Rate. An AR model of order p is defined as:

$$X(t) = c + \sum_{i=1}^p \varphi_i X(t-i) + \varepsilon(t) \quad (3.5)$$

Where $X(\cdot)$ is the process value at time \cdot , c is a constant, φ_i are the p parameters of the model and $\varepsilon(t)$ is the white noise used as the model input. p can be optimized as to minimize the prediction error (i.e. the error the model commits at time t with respect to the actual value of the real time series), for example using the Aikake Information Criterion (AIC), defined as

$$AIC = 2p - 2\ln(L) \quad (3.6)$$

Where L is the maximum value of the likelihood function and p is the number of estimated parameters. Given a set of models, AIC rewards goodness of fit (as assessed by the likelihood function), but also includes a penalty that is an increasing function of the number of estimated parameters, so that the smallest order that minimizes prediction error is used in the AR model. In fact, the order corresponding to the minimum value of the AIC function is chosen.

A bivariate AR model, with estimated spectra created from the test EEG recording X of unknown evoked potential frequency and the training recording Y with known frequency, can be used to compute coherence, defined as:

$$C_{X,Y} = \frac{|S_{X,Y}|^2}{S_{X,X}S_{Y,Y}} \quad (3.7)$$

So that $0 \leq C_{X,Y} \leq 1$. $S_{X,Y}$ is the cross-spectral density between X and Y, $S_{X,X}$ and $S_{Y,Y}$ are the auto-spectral densities for X and Y, respectively. The coherence is a normalization of the cross-spectral density (suitable for comparison between different combinations of test and train sets) and is maximized for similar spectra, i.e. for similar frequency information content. This approach only requires one electrode (selected as the one containing most information).

All the features described in this chapter were implemented and their discriminative power compared, as explained in the next chapter 4.

Chapter 4

Implementation

4.1 Training

Considering that the frequency of each letter of the speller is known, as is the symbol corresponding to each recording, the data set is divided in time sequences of 4 seconds of length each. The windowing was chosen because in literature offline training gives best results for time frames with duration of 4 seconds or more [6]: after 4 seconds in fact the improvement in classification accuracy is of negligible effect with respect to the lower ITR. However, as it will be detailed in chapter 5, during this project the change in accuracy relative to different window lengths was studied and its results were consistent with literature.

A training set was created from recordings on three subjects (two female and one male) who had high SNR: they were recorded using the single-stimulus paradigm described in the Material and Methods chapter 3, so that there would not be any additional stimuli in their field of vision that might cause lower attention from the subject. All frequencies used on the other subjects were recorded and the resulting high quality 10-second recordings were averaged across the three subjects and then processed like the test set: this procedure highlighted the desired frequency component and lowered any unwanted noise and distortions at frequencies that might be equivocated for the target stimulus one. Finally, the first four seconds of the averaged resulting time series were used to compute their correlation and coherence with the test recordings.

4.2 Features extraction and classification

4.2.1 Correlation

All 24 recordings (one for each blinking stimulus at one of the 6 chosen frequencies) were processed through a feature extraction algorithm and then a classification rule based on maximum cross-correlation.

The target frequency is extracted using two classifiers in series: first a 1024-point spectral estimate is computed using a monovariate AR model, with order estimated using AIC, its peaks isolated and the ones falling within the alpha rhythm range (9÷12 Hz) ignored. The remaining peaks close to the interface stimulation frequencies (within ± 0.4 Hz from the blinking frequency, to account for experimentally observed peak shifts) are taken into account: the three highest ones are considered as candidate classes for the following step of the classification, i.e. correlation on CCA-filtered data sets. This is useful to make the following analysis less computationally heavy, therefore shortening the processing time (necessary to obtain higher ITR in free spelling online experiments). With this approach, thresholding is unnecessary, meaning that no experimental user-dependent threshold has to be found beforehand (and possibly adjusted across trials). However, a different issue arises: as seen in literature [36], the “Midas Touch” problem (also known as the Zero Class problem) pertains to the incorrect classification of EEG recordings that do not reflect any evoked potential activity, while the subject is for example resting during the task or taking time to compose a chosen word. This is an important problem to consider for clinical applications, as the user might have finished communicating and random letters could be added to the already completed spelled word or, more importantly, incorrect letters might be added during communication, making the spelled word incomprehensible. To partially solve this problem, only peaks higher than the background noise (intended as the average power in the band 22÷25 Hz, which is not interested by the SSVEP in this study) are considered: if none are present, the algorithm classifies the recording as “rest”, i.e. no SSVEP is present. Furthermore, if only one possible candidate stimulation frequency is selected with this approach (i.e., there is no ambiguity in SSVEP frequency component selection), further classification is not needed and the recording is assigned to that class, decreasing processing time.

After the candidate classes are selected the second classification step consists of an analysis of similarity between a test and a training time series. The CCA necessary for the feature extraction process was implemented using the Matlab function `canoncorr()`, which computes the sample canonical coefficients for the n-by-d1 and n-by-d2 input data matrices. These matrices must have the same number of observations (rows) but can have different numbers of canonical variables (columns). The resulting canonical coefficients are organized in matrices, where the rows are the coefficients for the d1 or d2 input variables and the columns are the independent coefficient vectors. The coefficients are

then applied to each variable (i.e. the channels), so that the information is combined in a single fictitious channel, reducing the dimensionality but maximizing the information contained, since each channel is weighted according to their informative content by the CCA coefficients. The Matlab function creates several coefficient vectors, independent from each other, so that the first column creates the maximal correlation between the two data matrices and the following ones, if applied to the same matrices, only explain the remaining information. Therefore, the maximum amount of information is contained in the first column (see appendix A for more details) and only the first coefficient vector is applied to the data to be used for feature extraction. Afterward, the coefficients are applied to pairs of data matrices (as seen in Figures 4.1 and 4.2) and the Pearson coefficient is computed for the two resulting data vectors (of length n), using the Matlab function `corrcoef()`. The CCA is performed pair by pair on the reference sine-cosine signals, the averaged train set and the test set. The last two have already been acquired and preprocessed during training (see previous section), while the reference sine-cosine signals, commonly referred as Y_k , are completely created in Matlab as:

$$Y_k = \begin{pmatrix} \sin(2\pi f_k n) \\ \cos(2\pi f_k n) \\ \sin(4\pi f_k n) \\ \cos(4\pi f_k n) \end{pmatrix} \quad (4.1)$$

Where f_k is the selected flickering frequency from the typing matrix and n the discrete time: the length of the reference signals, i.e. n , must be the same as the length of the train set and test set, in this case 4 seconds, as required by CCA. The canonical variables are instead linked to the four different sinusoids (in contrast with the EEG data, where the variables are the different channels recorded), which are the sine and cosine wave at the selected flickering frequency and its harmonics: the number of harmonics was selected based on literature [20, 21] and experimental observation of the decrease in power of the harmonics in real SSVEP recordings. The sine and cosine were both included as the Pearson correlation index is the value of the cross-correlation function for a time delay equal to zero, therefore if two sine waves at the same frequency have a phase delay, their correlation coefficient will be low, although they have the same information content. Conversely, if different sinusoids with the same period are included, they will be weighted accordingly so that the one in phase with the other signal (i.e. the one that maximizes the correlation coefficient) will have a higher coefficient and therefore higher weight in the combined time series.

Afterward, the test set (of unknown frequency) is cross-correlated with the averaged train set at a specific SSVEP frequency and with the ideal sinusoids Y_k at the same frequency, varying the canonical coefficients applied to the two sets. Four Pearson indices

are computed, ρ_i :

$$\begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \rho_4 \end{pmatrix} = \begin{pmatrix} \rho(X^T W_X(XY) Y^T W_Y(XY)) \\ \rho(X^T W_X(X \hat{X}) \hat{X}^T W_X(X \hat{X})) \\ \rho(X^T W_X(XY) \hat{X}^T W_X(XY)) \\ \rho(X^T W_X(\hat{X}Y) \hat{X}^T W_X(\hat{X}Y)) \end{pmatrix} \quad (4.2)$$

Every ρ_i (i=1, 2, 3 or 4) is computed as the cross-correlation coefficient between the two spatially filtered data sets; X is the test set of unknown frequency, Y and \hat{X} are respectively the ideal sine-cosine matrix and the train set of the same frequency. $W_{\bullet}(XY)$ are the canonical coefficient vectors created between the test set and the ideal sinusoidal waves for all stimulation frequencies, $W_{\bullet}(\hat{X}Y)$ are between the train set and the ideal sinusoidal waves of the same frequency and $W_{\bullet}(X\hat{X})$ are between the test set and the train set, for all candidates. The values of each ρ_i is stored for each candidate frequency and then the $\hat{\rho}$ is computed, with the formula defined in the previous chapter 3 (i.e. as the sum of the squared ρ_i). The maximum value in the resulting vector of $\hat{\rho}$ indicates which frequency is the target one.

Each coefficient ρ_i was also evaluated, to establish which features were the most discriminative, so that the algorithm processing time could be further improved by reducing the number of spatial filters to calculate and apply (see chapter 5).

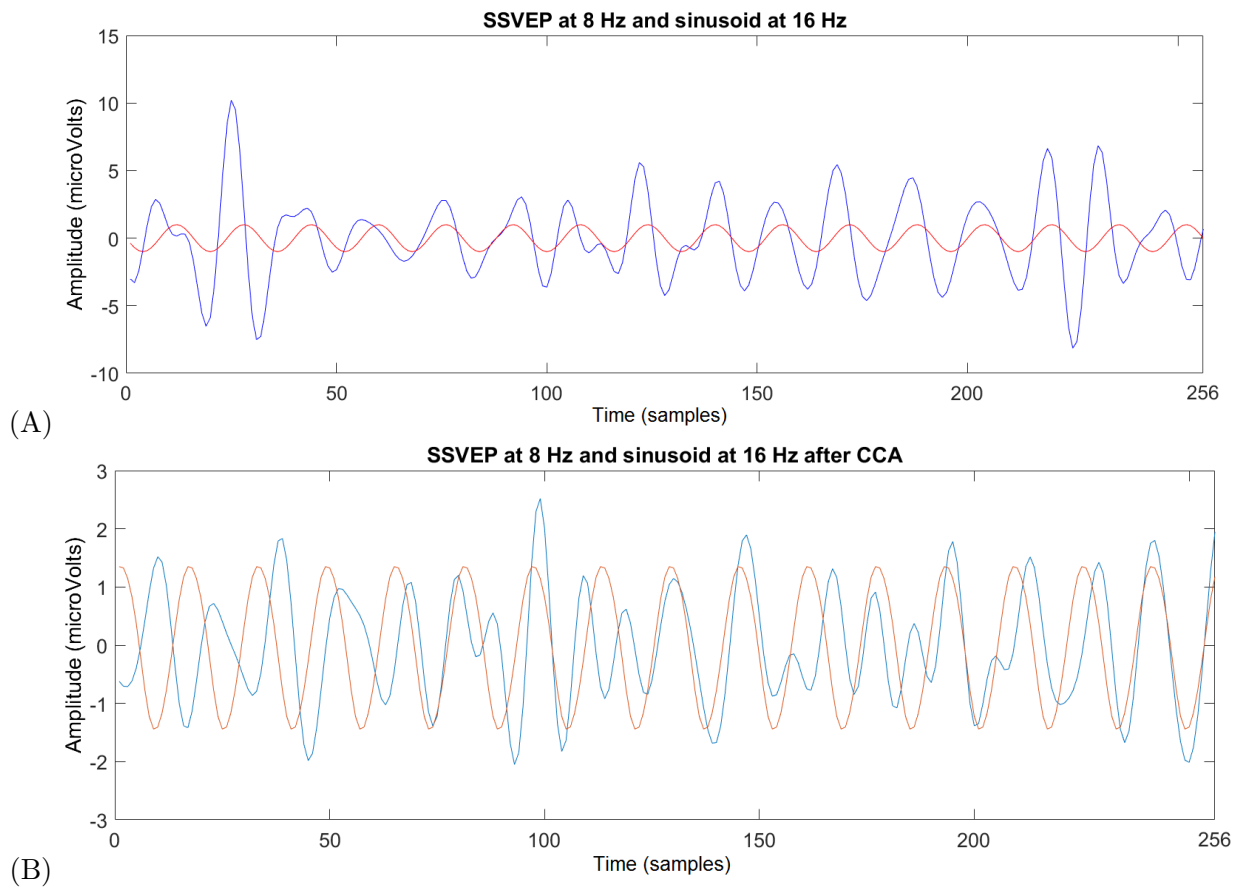


Figure 4.1: Effect of CCA: overlapped real EEG and ideal sinusoid at same frequency before (A) and after (B) CCA. Note that similar frequency content allows for a good overlap of the signals, therefore their cross-correlation at zero time lag is increased.

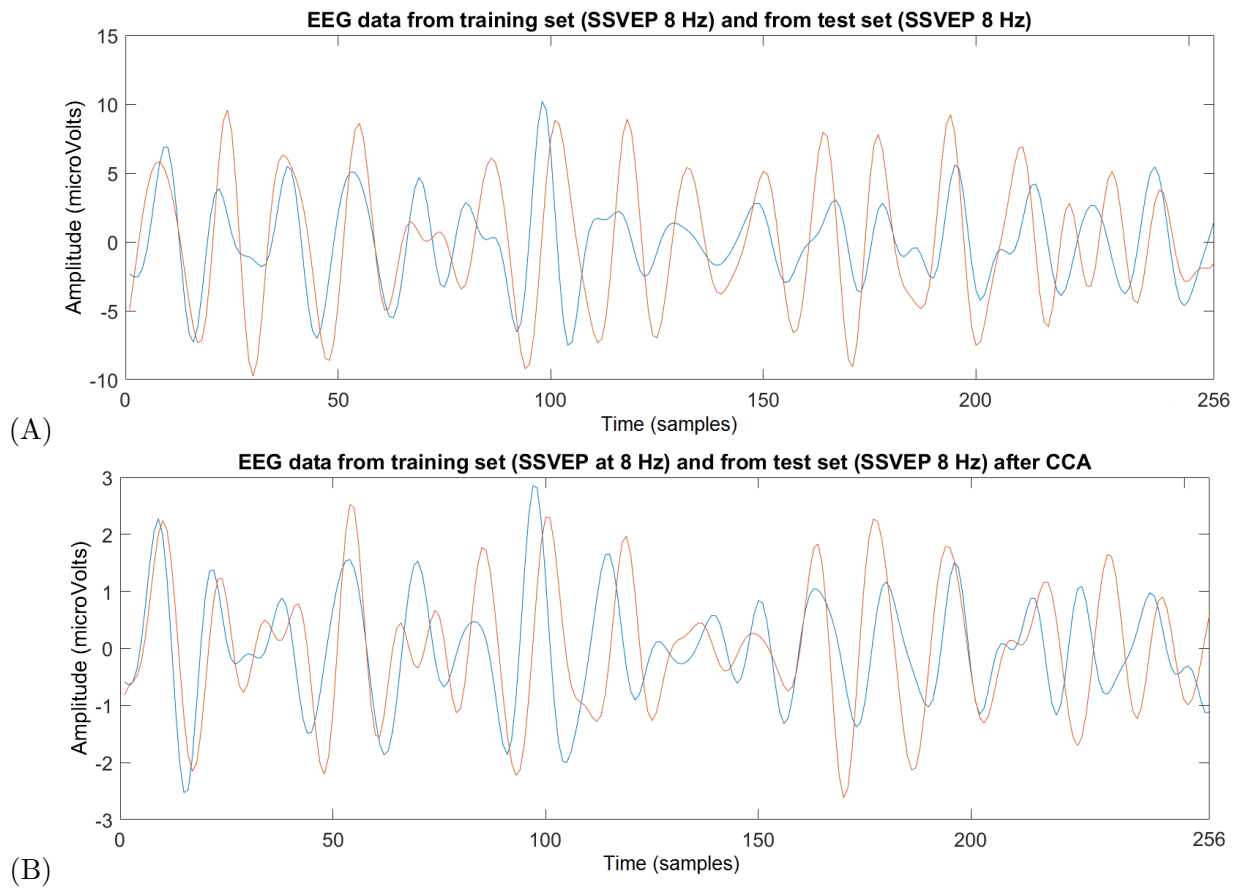


Figure 4.2: Effect of CCA: overlapped real test EEG and train EEG at same frequency before (A) and after (B) CCA. Note that similar frequency content allows for a good overlap of the signals, therefore their cross-correlation at zero time lag is increased.

4.2.2 Coherence

Computing the coherence (i.e., normalized cross-spectrum) between temporal series is useful to identify significant frequency-domain correlation at specific frequencies. The difference between temporal correlation and coherence is that, unlike time domain operations, the frequency domain is not affected by phase lags between the two series, allowing easier identification of specific recurring frequency components. In the context of this study, after a selection of the three most likely classification candidates as previously discussed, the training set at a known frequency and the test set not yet classified are used as input time series for a bivariate autoregressive analysis. To reduce computational complexity, the data set is also undersampled at a fixed rate, using the Matlab function `resample()`, in this case from 256 Hz to 64 Hz sampling frequency. This value has been chosen so that the updated Nyquist frequency (32 Hz) would still be well above the maximum signal frequency to detect (21 Hz).

After the undersampling operation, a bivariate autoregressive model is estimated, using a model order p between 9 and 12, optimized using AIC. The range was defined based on experimental values of order often used in EEG modeling and the prediction error criterion chosen based on literature on EEG modelization [47]: only few features (i.e. frequency components) are relevant to the present study and therefore a complex model is not required. In order to completely describe the bivariate model where the EEG train and test sets are respectively the variables x and y , four coefficient vectors are necessary: a_{xx} , a_{yy} , a_{xy} , a_{yx} . To completely describe the process, it is also necessary to calculate the auto- and cross-correlation between its input white noises, given the hypothesis that each of them is an additive Gaussian white noise with zero mean. With this premise, the p th-order model calculated can be defined as:

$$\begin{aligned}
 \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} &= \begin{pmatrix} x(t-1) \\ y(t-1) \end{pmatrix} * A^I + \begin{pmatrix} x(t-2) \\ y(t-2) \end{pmatrix} * A^{II} + \begin{pmatrix} x(t-p) \\ y(t-p) \end{pmatrix} * A^{(p)} + E = \\
 &= \begin{pmatrix} x(t-1) \\ y(t-1) \end{pmatrix} \begin{pmatrix} a_{xx}(1) & a_{xy}(1) \\ a_{yx}(1) & a_{yy}(1) \end{pmatrix} + \begin{pmatrix} x(t-2) \\ y(t-2) \end{pmatrix} \begin{pmatrix} a_{xx}(2) & a_{xy}(2) \\ a_{yx}(2) & a_{yy}(2) \end{pmatrix} + \dots + \\
 &\quad + \begin{pmatrix} x(t-p) \\ y(t-p) \end{pmatrix} \begin{pmatrix} a_{xx}(p) & a_{xy}(p) \\ a_{yx}(p) & a_{yy}(p) \end{pmatrix} + \begin{pmatrix} \epsilon_x(t) \\ \epsilon_y(t) \end{pmatrix}
 \end{aligned} \tag{4.3}$$

where E is the vector containing the input white noises, $x(\cdot)$ and $y(\cdot)$ the input time series values at time \cdot and A is a matrix composed of the four coefficients, which are defined as vectors, including their constant terms:

$$a_{xx} = \begin{pmatrix} 1 \\ a_{xx}(1) \\ a_{xx}(2) \\ \dots \\ a_{xx}(p) \end{pmatrix} \quad (4.4)$$

$$a_{yy} = \begin{pmatrix} 1 \\ a_{yy}(1) \\ a_{yy}(2) \\ \dots \\ a_{yy}(p) \end{pmatrix} \quad (4.5)$$

$$a_{xy} = \begin{pmatrix} 0 \\ a_{xy}(1) \\ a_{xy}(2) \\ \dots \\ a_{xy}(p) \end{pmatrix} \quad (4.6)$$

$$a_{yx} = \begin{pmatrix} 0 \\ a_{yx}(1) \\ a_{yx}(2) \\ \dots \\ a_{yx}(p) \end{pmatrix} \quad (4.7)$$

Each of these vectors value is calculated using the Yule-Walker equations: the algorithm first estimates the first p samples of the correlation function (with biased normalization) between each pair of x and y , using the Matlab function `xcorr()` and then computes the matrix R and the vector r , using the auto- and cross-correlation values previously calculated, defined as:

$$R = \begin{pmatrix} R(0) & R(-1) & \dots & R(-p+1) \\ R(1) & R(0) & \dots & R(-p+2) \\ \dots & \dots & \dots & \dots \\ R(p-1) & R(p-2) & \dots & R(0) \end{pmatrix} \quad (4.8)$$

$$r = - \begin{pmatrix} R(1) \\ R(2) \\ \dots \\ R(p) \end{pmatrix} \quad (4.9)$$

So that the coefficient vectors are easily calculated as the product of the inverse of R and r . Afterward, the algorithm can proceed with the computation of the auto- and cross-spectral estimates for the two time series (seen in Figures 4.3 and 4.4), according to well-tested methodologies found in literature [48]:

$$S = N * Pvar * N' \quad (4.10)$$

$$N = inv(I - C) \quad (4.11)$$

Where S is the spectral estimate, N' is the transpose of N and $Pvar$ is the matrix containing the input white noises variances and covariances. I is the identity matrix and

C is the Fourier transform of the sum of the model coefficients, i.e.:

$$C = \begin{pmatrix} \sum_{k=1}^p a_{xx}(k) * e^{i*k*\omega} & \sum_{k=1}^p a_{xy}(k) * e^{i*k*\omega} \\ \sum_{k=1}^p a_{yx}(k) * e^{i*k*\omega} & \sum_{k=1}^p a_{yy}(k) * e^{i*k*\omega} \end{pmatrix} \quad (4.12)$$

After the cross-spectrum is estimated, its normalized value, i.e. the coherence, is used as the classification feature, because absolute spectral values could vastly change across recordings, compromising the ability of the classifier to identify the correct target class. The value of the coherence peak closest to the stimulation frequency of the train set in exam (within the range of ± 0.4 Hz from the exact value) is evaluated. Then the maximum out of these three values is taken as the one corresponding to the target class. In fact, the coherence should be highest (i.e. closest to 1), when the two time series both have peaks in their real spectra, which are modeled by the AR spectral estimates. The other advantage of using these estimates is that because of the order chosen in this algorithm, they also have less peaks and are generally smoother: less distortion components are taken into account, theoretically decreasing the misclassification error.

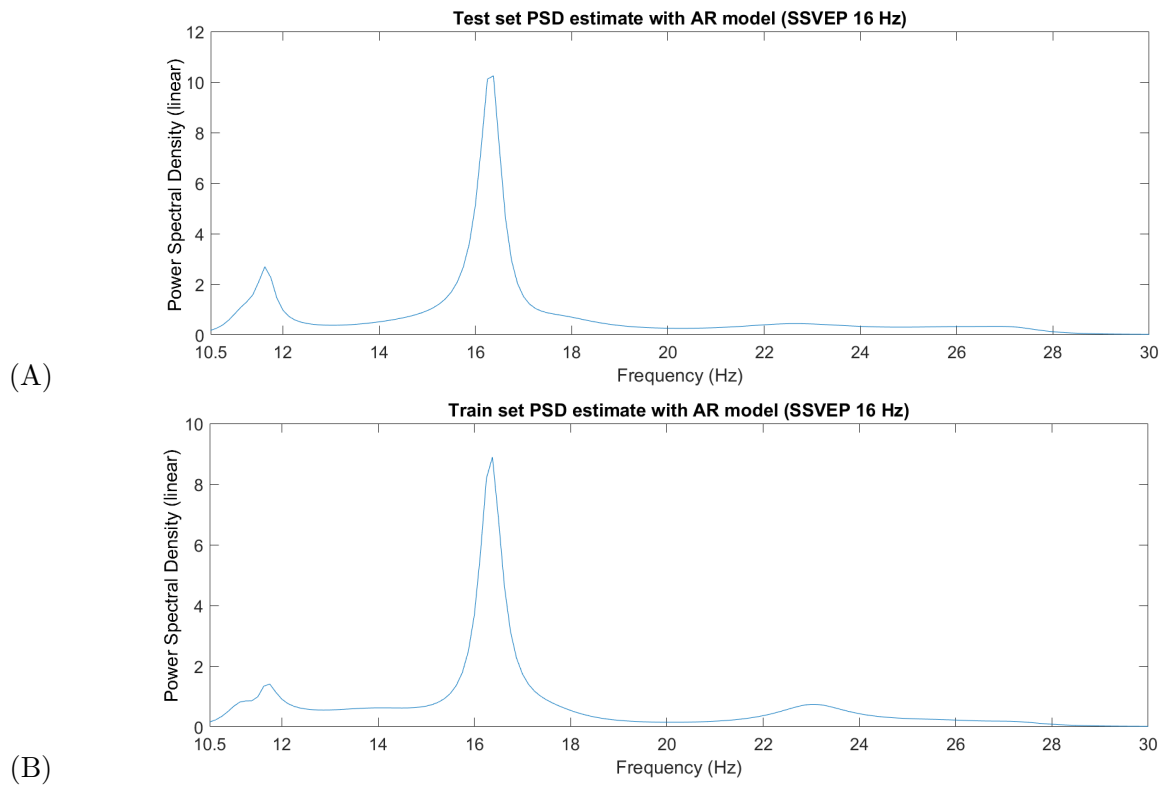


Figure 4.3: Spectra of the test (A) and train (B) recordings, estimated with AR model. The stimulus frequency is 8 Hz for both: note the peak at its second harmonics, 16 Hz.

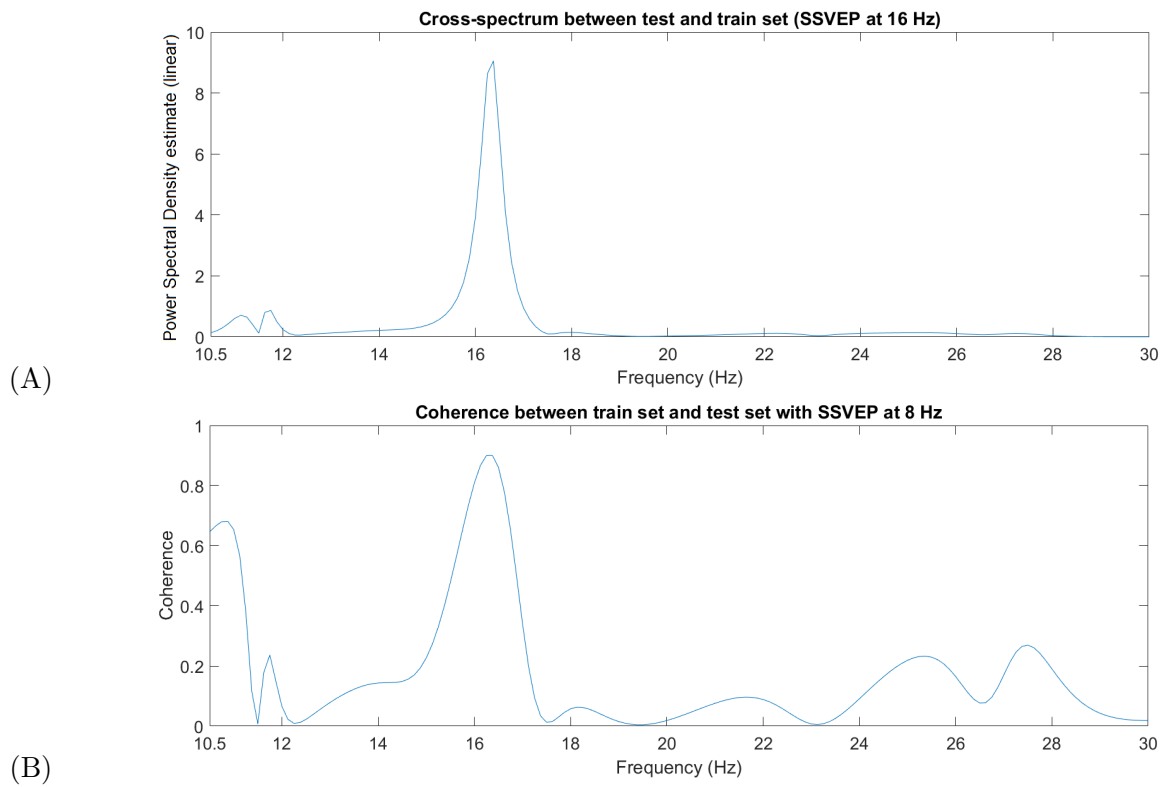


Figure 4.4: Cross-spectrum (A) and coherence (B) between the train and test recording estimated with AR model. The stimulus frequency is 8 Hz for both: note the peak at its second harmonics, 16 Hz.

Chapter 5

Results

5.1 Simulated data

The two frequency classifiers were tested on ten different simulations (created as described in the chapter 3) of the seven possible stimulation frequencies. Each times series was divided in four-second long windows and classification was carried out on the channels with highest SSVEP quality (defined as the ones with highest signal amplitude with respect to the distortions, as discussed in chapter 3). The resulting median accuracy in frequency classification was $95.2 \pm 2.6\%$ when using correlation features and $85.7 \pm 4.3\%$ for coherence. From these results (which are reported in Tables 5.1 and 5.2, it is apparent that a correlation-based approach is slightly more effective than a coherence-based classification. In particular, coherence is more sensitive to alpha rhythm components and other disturbances, especially when they appeared at less than 1 Hz to the SSVEP component. In fact, such components close to each other in frequency may cause a slight shift in the estimated spectrum peaks, lowering coherence. Furthermore, increasing the standard deviation of the white noise or the amplitude of the distortion components did not cause additional misclassifications until the SNR reached values inferior to 1. Both approaches were found to be robust to white noise, while misclassifications were mostly caused by unwanted frequency components with amplitudes close to the modeled SSVEP response. However, if both training and test set had similar frequency content (including in terms of distortions), classification errors were less frequent.

It is also important to note that, when a phase delay was introduced for the SSVEP signal (i.e. only on the sinusoidal component at the target frequency), the calculated Pearson index decreased. When the test and train signals were out of phase, more classification errors were introduced for the correlation-based approach, with median accuracy slightly decreasing ($90.48 \pm 4.2\%$). Conversely, coherence is a more robust method to phase shifts, as these simulations demonstrated: the accuracy did not decrease when phase shifts were introduced. This is due to the fact that coherence is a measure

of similarity between spectral magnitudes and it is often used to compare time series with similar frequency content, regardless of their relative phase delay.

Frequency	8	9	13	15	17	19	21
8	1	0	0	0	0	0	0
9	0	0.92	0	0.04	0.04	0	0
13	0	0	0.96	0	0.04	0	0
15	0.04	0	0	0.96		0	0
17	0	0.13	0	0	0.87	0	0
19	0	0	0	0	0	1	0
21	0.04	0	0	0	0	0	0.96

Table 5.1: Normalized confusion matrix of the correlation approach computed across ten simulations: the row indicates target class; the column is the assigned class. Total accuracy is 95.2 ± 5.6 %.

Frequency	8	9	13	15	17	19	21
8	1	0	0	0	0	0	0
9	0.08	0.80	0	0	0.04	0.08	0
13	0	0.04	0.84	0.04	0	0.04	0.04
15	0.08	0	0	0.92	0	0	0
17	0	0.08	0.04	0	0.88	0	0
19	0	0	0	0.17	0.04	0.71	0.08
21	0.04	0	0.04	0.08	0	0	0.84

Table 5.2: Normalized confusion matrix of the coherence approach computed across ten simulations: the row indicates target class; the column is the assigned class. Total accuracy is 85.7 ± 7.3 %.

5.2 Experimental data

Out of the 17 subjects recruited, the recordings from ten of them were used to test the discrimination power of the extracted features: the other subjects (five male and two female) were discarded either because incorrect application of the electrode cap caused fluctuation of the ground electrode, which in turn compromised the quality of the recordings, or because even in a single-stimulus paradigm they were unable to produce significant SSVEP response. In one case a subject produced a strong alpha rhythm even during the concentration task, causing systematic classification errors, consistent with the simulation results (refer to previous section on the matter).

Out of all the subjects, there were not strong differences in terms of demographics: all of them were in the same age range (every subject was between the age of 20 and 29), had university-level education and none of them had ever received diagnosis of mental or chronic illness. Only two of them had previous experience with BCIs (which could theoretically account for a better ability to perform the task required) and all candidates met the criteria necessary for SSVEP studies (no previous history of epilepsy, recurring migraines or photosensitivity). No subject was color blind and while several of them had a vision condition, all of them wore prescription lenses that corrected their vision to normal. These characteristics made the subjects ideal candidates to test the interface, since better performances have been reported in younger subjects; however, it did not allow for more in-depth demographics analysis.

More heterogeneity was found in regard to subject habits: specifically, use of psychoactive substances, weekly hours spent on a computer and sleeping habits (see Figures 5.1 and 5.2). In particular, 15 of the 17 subjects had drunk coffee in the last 24 hours: the remaining two found the task more tiring than average. Of course, this result does not attain statistical relevance because of the small sample size; it is however possible that people who had consumed caffeine felt more awake and therefore inclined toward concentration tasks. On the other hand, no correlation was found between alcohol consumption (which had happened the evening before for three subjects), recent nicotine consumption (four subjects) or antihistamine use (two allergic subjects). Sixteen subjects had had at least 6 hours of sleep and were reasonably well-rested; however, the two best performances were both from people who had slept for longer (7 and 8 hours) and reported not being tired: fatigue is certainly a factor that causes a decrease in focus. Regarding computer usage, those subjects who spent a significantly shorter time on computer in general requested more breaks during the experiment and were more likely to report dry eyes toward the end of the task. This is expected as intense lights can be annoying for certain subjects, although practice with the interface would lessen the issue for most of them. In regard to eye fatigue, an experiment was also conducted on 5 subjects, who were asked to try focusing on a stimulus displayed on a CRT monitor first and then on the same stimulus on a LCD monitor: all of them both

had worse performances and found the task more difficult and tiring.

Five of the subjects also regularly played videogames, but no relevant performance improvement was recorded with respect to other subjects; however, all of them reported minimal difficulties in focusing on the flickers, regardless of other visual stimuli in their field of vision, while some of the subjects who did not play videogames were more annoyed and distracted by them. This could be explained by the relative similarities between the concentration task of the experiment and videogame mechanics and is consistent with literature [36].

The alpha rhythm of the subjects was also studied, because its suppression during an open-eye task is not always complete and has to be taken into account for both paradigm development and signal processing. The subjects were asked to close their eyes and relax for 20 seconds, both before and after the concentration task: the results were that the dominant alpha frequency (i.e. the one with maximum power even among those in the alpha band) remained almost constant, with shifts of less than 0.5 Hz in 88% of cases. Further studies on inter-trial stability revealed that the same subjects repeating the experiment, even one or two months after the previous recording, had similar alpha wave patterns, concentrated around similar frequencies. All subjects had alpha activity concentrated in the range between 8.8 and 12.4 Hz, with average main alpha frequency at 11 ± 0.7 Hz for females and $10.6 \text{ Hz} \pm 0.9$ Hz for males. Generally speaking, women exhibited higher frequency and higher amplitude alpha rhythms (in one specific case, high enough that even with open eyes its suppression was not enough to detect evoked potentials). However, some men exhibited high frequency alpha waves, close to 12 Hz. Considering the stability and inter-subject variability of this rhythm, these results should be taken into account for personalization of the paradigm (see chapter 6).

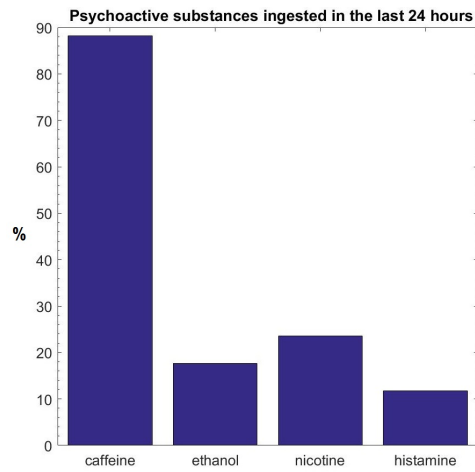


Figure 5.1: Questionnaire results on psychoactive substance ingestion by the test subjects.

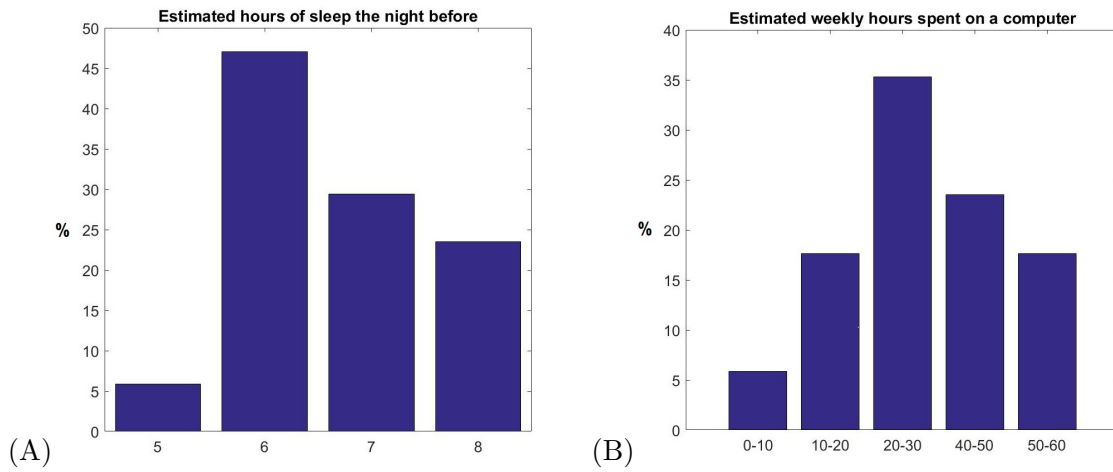


Figure 5.2: Questionnaire results on sleeping habits of the test subjects (A) and their estimate of weekly hours spent on the computer (B).

To evaluate classifier performances in BCIs, results are usually organized in a structure known as confusion matrix [49], a n-by-n matrix in which the (i,j)-th entry represents the number of samples from a class i classified as belonging to a class j. A confusion matrix provides exhaustive information on how each class is classified by the BCI system. Confusion matrices were calculated for each subject, in order to study the system performances and discuss median results (experimental results obtained during this project are reported in Tables 5.3 and 5.4). In particular, an important index is the classifier accuracy, i.e. number of correctly classified recordings out of the total number of recordings considered for the single subject. Another criterion often used in BCI studies is the Cohen's Kappa Coefficient, κ , which considers chance agreements, i.e. compares the classifier performance with a random classifier and as such is regarded as a more robust measure of accuracy:

$$\begin{aligned} p_e &= \sum_{i=1}^M \frac{n_i m_i}{N^2} \\ \kappa &= \frac{p_o - p_e}{1 - p_e} \end{aligned} \quad (5.1)$$

Where p_o is the accuracy as previously defined, n_i is the sum of the recordings classified in the i-th class, m_i is the sum of the recordings belonging to the i-th class and N is the total number of samples. Negative κ indicates performances worse than a random classifier (in case of systematic classification errors), null κ is the value of a random classifier and coefficient equal to 1 is the perfect classifier.

Frequency	8	13	15	17	19	21	Total
8	10	1	1	0	0	0	12
13	0	12	0	0	0	0	12
15	0	0	12	0	0	0	12
17	4	1	0	7	0	0	12
19	1	2	0	0	8	1	12
21	0	1	2	1	0	8	12
Total	15	17	15	8	8	9	72

Table 5.3: Confusion matrix of the correlation approach computed for one subject (male, age 23): the row indicates target class; the column is the assigned class. Total accuracy 79.17%.

Frequency	8	13	15	17	19	21	Total
8	10	1	0	1	0	0	12
13	0	11	1	0	0	0	12
15	0	0	12	0	0	0	12
17	4	1	0	7	0	0	12
19	1	2	0	0	9	0	12
21	0	2	4	1	0	5	12
Total	15	17	17	9	9	5	72

Table 5.4: Confusion matrix of the coherence approach computed for one subject (male, age 23): the row indicates target class; the column is the assigned class. Total accuracy 75%.

Classification accuracy can vary greatly based on the number of electrodes used (for a maximum of 8, covering the entirety of the visual cortex) and the time window length (for a maximum of 5 seconds according to literature [50], after which no improvement is seen). As such, both these factors were considered for the ten best candidates, in order to determine whether or not the algorithm should be personalized for these two parameters. While the algorithm could be customized for each candidate, the added training time would be impractical, especially if there is not a remarkable improvement in accuracy. When considering the number of electrodes used for feature extraction, it can be noted that, if the same evoked response is present in more than one channel, this component will be enhanced if more electrodes are considered, while random fluctuations and noise stemming from a small region of the visual cortex will be damped. However, the

SSVEP response is maximally produced by a specific area and hemisphere of the central nervous system, varying from subject to subject but constant across trials. Therefore, a lower number of electrodes could be used, allowing for a faster mounting of the EEG cap, increasing feature extraction speed and solving the curse of dimensionality issue described in chapter 2. Electrodes have to be chosen carefully, because not all channels of the occipital and parieto-occipital region respond in the same way to the visual stimulus: in certain cases the evoked potential is completely absent or there is strong peaking at a different frequency (e.g., an unsuppressed alpha rhythm component). Furthermore, sometimes electrodes might drift or have lower SNR because of incorrect application of the electrode gel. As such, it is important not to include these electrodes in the feature extraction process for a more robust classification. The electrodes were selected according to the methodology described in chapter 3 and ordered in terms of SNR, then the classification was carried out with an increasing number of electrodes, following this order. In the correlation approach, higher dimensionality truly improves accuracy at first, causing a peak at 3 channels, but for higher numbers of electrodes accuracy slightly decreases and then stabilizes to a plateau value, as shown in Figure 5.3 and Table 5.5. All subjects considered had best classification accuracy at either 3, 4 or 5 electrodes and when 4 or 5 channels caused an improvement, it was not very significant (less than 2% in all cases). It can be then inferred that 3 accurately chosen electrodes could be potentially sufficient for the setup of this system. On the other hand, for the frequency features, an averaging of the coherence values was attempted across channels, but it did not cause any improvements in the classifier efficacy. In addition, it corresponded to longer processing times, because of the higher computational power required to formulate several independent AR models (one for each pair of corresponding channels in the train and test sets). As such, one-channel only computation was considered in the following analyses.

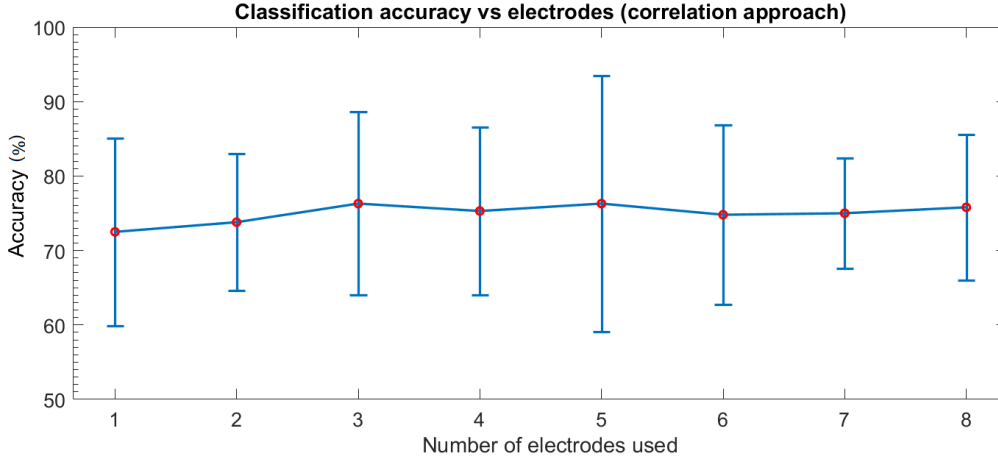


Figure 5.3: Classification accuracy for the correlation approach across ten subjects with relation to number of electrodes used (using four-second long time windows). Notice the peak at 3 electrodes and the lack of significant improvement with higher number of electrodes.

Number of electrodes	Accuracy (%)
1	72.5±12.6
2	73.8±9.2
3	76.3±12.3
4	75.3±11.3
5	76.3±17.2
6	74.8±12.1
7	75±7.4
8	75.8±9.8

Table 5.5: Percentage median accuracy and interquartile ranges for the correlation approach with respect to the number of electrodes used (using four-second long time windows).

Furthermore, a study on the most effective window length was carried out: the ideal length of the recording should be enough to easily discriminate the signal component from the background noise, but not tiring or frustrating for users, i.e. causing slow communication. Many literature studies identify these length between 1 and 5 seconds: for longer windows, the accuracy of most classifiers plateaus. As such, the classifiers tested on different time windows for ten subjects: each recording was 20 seconds long and the first N seconds (with N=1, 2, 3, 4 or 5) were used. The reason for this is that

the first few seconds would probably coincide with the time where the user was most focused on the task (as many subjects reported focus tended to drift toward the end of each longer recording). The results showed a remarkable uniformity: both classifiers performed best with 4-second long recordings, for all subjects. Longer windows did not show any improvement on the results and manual analysis of the spectral estimates did not show increased SNR, as seen in Figure 5.4 and Table 5.6.

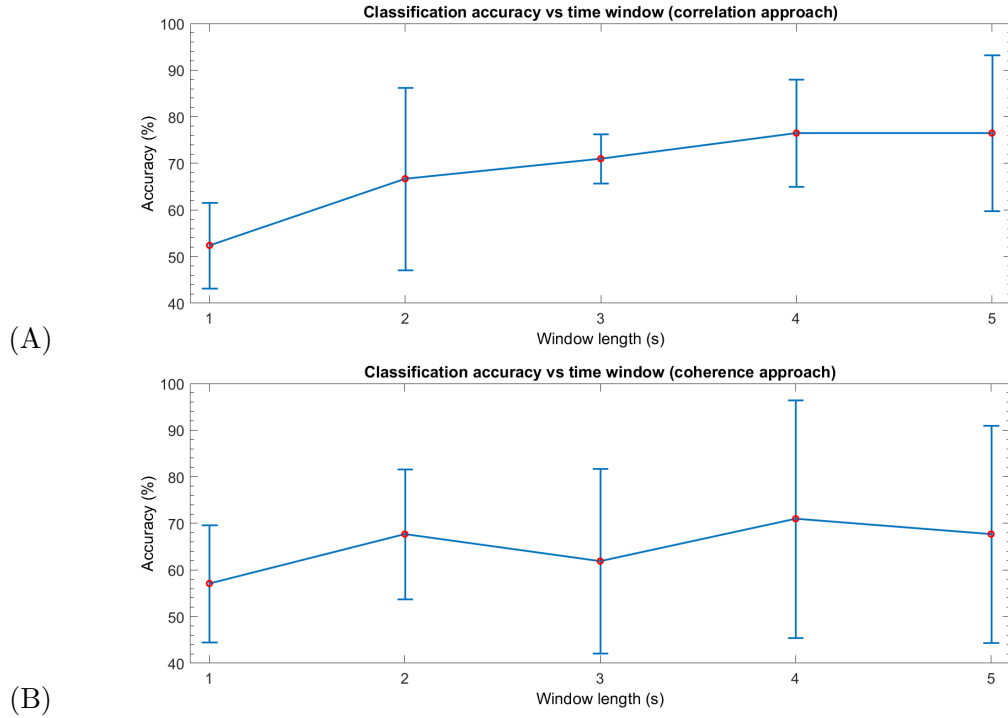


Figure 5.4: Classification accuracy for the correlation (three electrodes) and coherence (one electrode) approaches across ten subjects with relation to the window length used. Notice the lack of improvement for windows longer than 4 seconds.

Window length (s)	1	2	3	4	5
Correlation accuracy(%)	52.4±9.2	66.7±19.6	71±5.3	76.5±11.5	76.5±16.7
Coherence accuracy(%)	57.1±12.6	67.7±13.9	61.9±19.8	71±25.5	67.7±23.3

Table 5.6: Percentage median accuracies and interquartile ranges for the correlation (three electrodes) and coherence (one electrode) approaches with respect to the data length used.

With the fixed parameters of four-second long recordings and, respectively, three and

one channels, the correlation-based classifier showed a median accuracy of $76.5 \pm 11.5\%$ (maximum value: 93%, minimum value: 58.3%) and the coherence-based one of $71 \pm 25.5\%$ (maximum value: 88.9%, minimum value: 48.6%) and κ respectively equal to 0.72 and 0.43, marking both approaches as free from systematic errors and well above the performances of a random classifier. These results are in line with literature, where most approaches based on correlation between a test set and an ideal sinusoid or an averaged train set, attest results higher than 60% in almost all cases [36]. No comparison with literature could be done between the coherence approach and similar studies, as this feature has not been used in previous works on SSVEP-based BCIs, to the best of the author’s knowledge. The high interquartile ranges can be explained by the physiological high variability across subjects in terms of EEG signals and their features.

Lastly, each correlation feature (i.e. each ρ_i coefficient) was also separately evaluated in terms of discriminative power, so that the minimum number of features could be used, to improve computational times without compromising classification accuracy. Specifically, the previously defined ρ_3 (with CCA filters calculated between the test set and the ideal signals) and ρ_4 (with CCA filters calculated between the train set and the ideal signals) were found to be redundant and did not increase classification accuracy. As such, they were not employed in the final implementation of the classification algorithm: this is possibly due to the fact that the calculated filters did not result in a significantly different weighting of channels, combining information in a similar way to the first two (in particular considering that the data sets were the ones used in the computation of ρ_1).

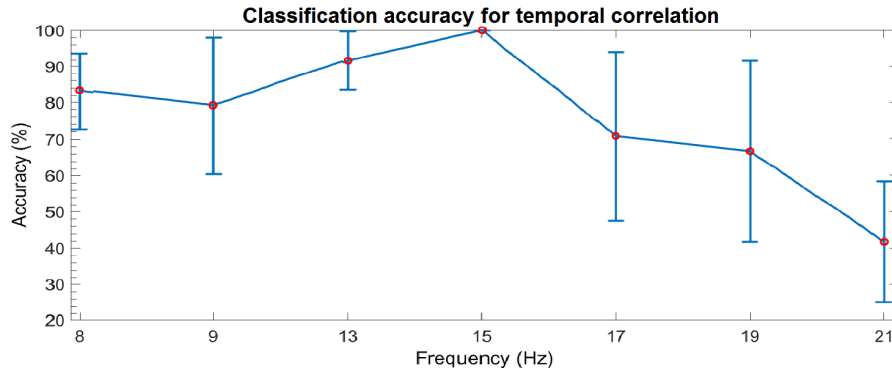


Figure 5.5: Classification accuracy for the correlation approach across ten subjects per single frequency. Note the decrease in accuracy at higher frequencies (higher than 15 Hz).

Frequency	8	9	13	15	17	19	21
Accuracy (%)	83.3±10.5	79.2±18.8	91.7±8	100±0	70.8±23.3	66.7±25	41.7±16.7

Table 5.7: Percentage median accuracy and interquartile ranges for the correlation approach with respect to the target frequency.

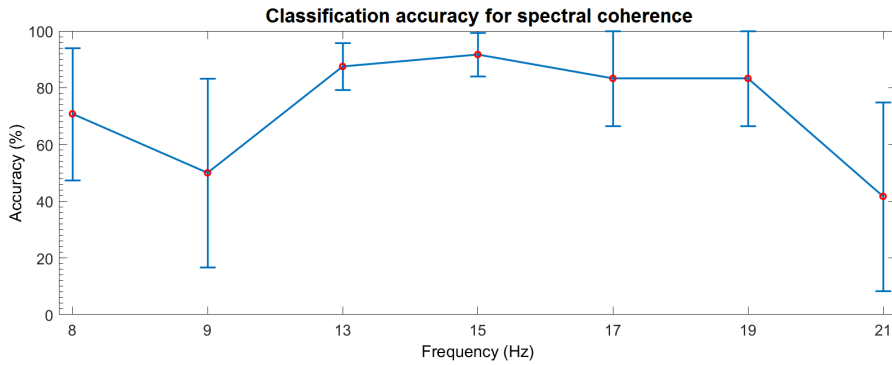


Figure 5.6: Classification accuracy for the coherence approach across ten subjects per single frequency. Note the higher accuracy at 17 and 19 Hz with respect to the correlation approach.

Frequency	8	9	13	15	17	19	21
Accuracy (%)	70.8±23.3	50±33.3	87.5±8.3	91.7±7.7	83.3±16.7	83.3±16.7	41.7±33.3

Table 5.8: Percentage median accuracy and interquartile ranges for the coherence approach with respect to the target frequency.

Analyses of peak shifts in the spectral estimates (always within the ± 0.4 Hz range previously mentioned), Pearson's indices and coherence value were conducted, as correlation and coherence between a test set and a train set with SSVEP at known frequency decrease when the main frequency components of the test recording are placed farther apart from the frequency candidate. This allowed to estimate that the minimum variation between stimulation frequencies should be 1 Hz, so that there is not an overlap between two different bands of evoked response: experimentally 1 Hz was the minimum variation that did not cause unwanted equivocation between neighbor classes. Interesting results can be obtained comparing the accuracy in classification frequency by frequency (as seen in Figures 5.5 and 5.6 and Tables 5.7 and 5.8), as all 10 subjects demonstrated similar patterns for the same stimulation frequency: 21 Hz stimuli were found less annoying by the subject but had low SNR, with spectral estimates much less smooth than at lower frequencies (see Figure 5.8 for details). In particular disturbances at frequencies in the band of interest were strong components of the final filtered EEG signal and accounted for most of the inaccuracies of the classifier. In fact if 21 Hz stimuli were omitted from the interface, an increase in accuracy was apparent: respectively, the correlation and coherence classifiers had median accuracy equal to $83.3\pm 13.6\%$ and $76.5\pm 21.2\%$. It can be observed that 13 Hz and 15 Hz SSVEP were consistently the most correctly classified commands for both algorithms, and while coherence was in general less discriminative than correlation, it did perform better for 17 and 19 Hz, both frequencies closer to the limit of the range normally considered for this type of stimulations ($8\div 22$ Hz). Another interesting result concerns the low-band frequencies: 8 and 9 Hz SSVEPs were not considered part of the signal band, since their classes were favored by the classifier because of the $1/f$ shape of EEG power, even when evoked potentials are not present. As such, their second harmonics (16 and 18 Hz) were considered instead and then elaborated and classified with the same methods as the other evoked potentials: this did not however particularly influence the accuracy of the classifier, as their amplitude was still remarkable (see Figure 5.7). It can be noted however that a small decrease in accuracy of the 9 Hz frequency is present, probably caused by a slight decrease in amplitude of the harmonic response.

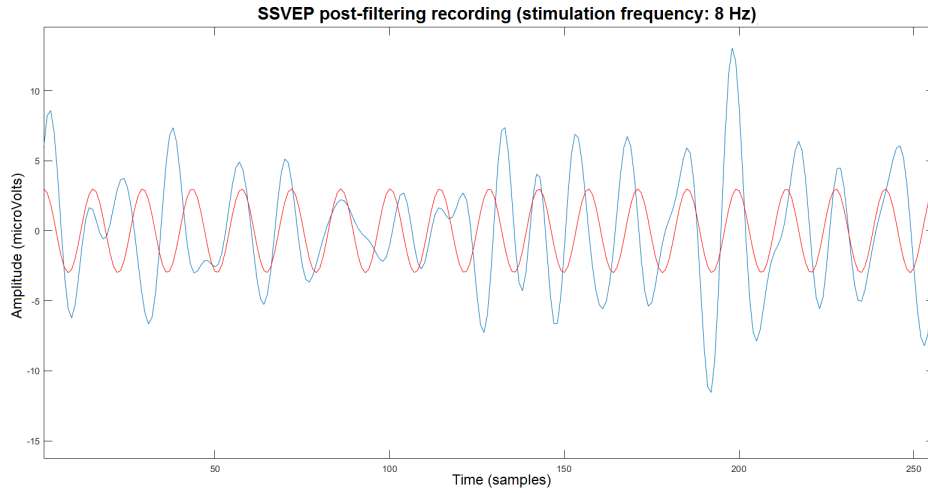


Figure 5.7: The overlapped recording of a subject under 8 Hz flashing stimulus and an sinusoid with frequency equal to 16 Hz. Note the strong second harmonic component of the SSVEP signal, validating its use in the classification algorithm.

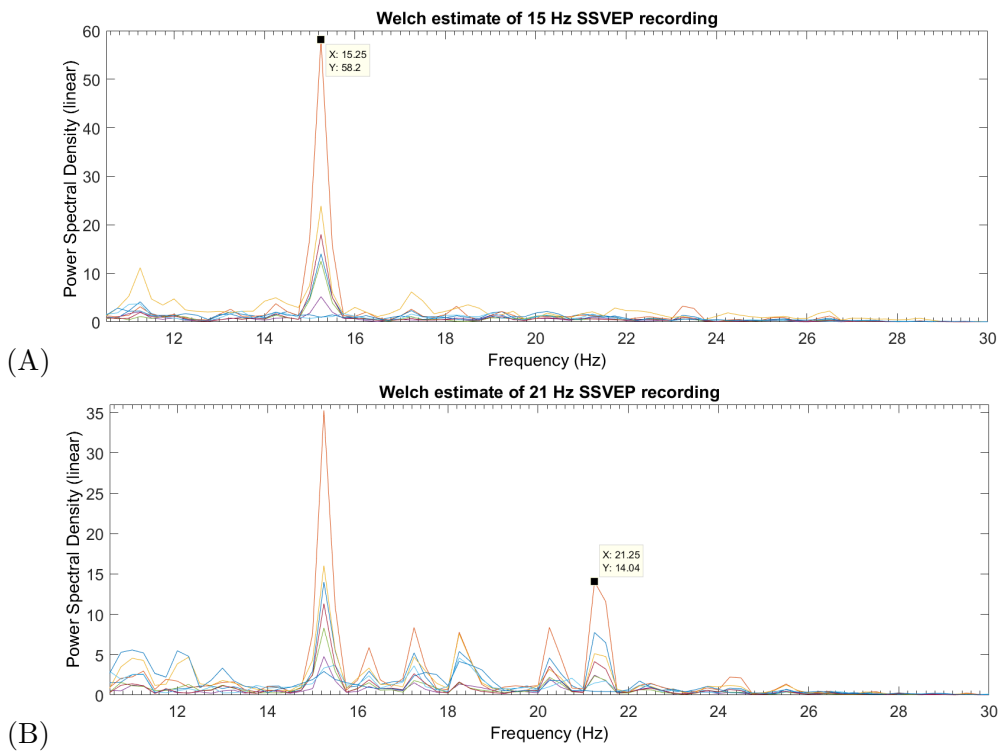


Figure 5.8: Welch estimate of 15 Hz SSVEP recording (A) and 21 Hz SSVEP (B) on the same subject across the eight electrodes recorded. Note the difference between the two spectra in terms of SNR.

Given the confusion matrix, it is possible to extrapolate the number of false positives, i.e. misclassified signals, which is a good measure to understand whether a certain class was preferred by the system in ambiguous cases. Given the definition of Specificity (Sp):

$$Sp = TN/(TN + FP) \quad (5.2)$$

Where TN stands for True Negatives and FP for False Positives. By considering the specificity, the classifiers can be separately evaluated in terms of FP rate per frequency, equal to 1-Sp, as seen in Table 5.9.

Frequency	8	9	13	15	17	19	21
Sp (correlation)	0.9306	0.9931	0.9444	0.8958	0.9583	1	1
1-Sp (correlation)	0.0694	0.0069	0.0556	0.1042	0.0417	0	0
Sp (coherence)	0.9375	0.875	0.9236	0.9306	0.9444	0.9792	0.9931
1-Sp (coherence)	0.0625	0.125	0.0764	0.0694	0.0556	0.0208	0.0069

Table 5.9: Specificity and FP rate per frequency, to evaluate the nature of the classification errors.

It is apparent that higher frequencies have higher specificity, because test data is rarely misclassified as high frequency. This might be due to the fact that higher frequencies not only tend to have lower signal power (caused by the $1/f$ trend of the EEG spectrum), but their noise is lower too. On the other hand, coherence seems to overestimate the 18 Hz component of most recordings, while correlation has satisfying specificity at all frequencies, with a small decrease at 15 Hz.

The Matlab algorithms that yielded these results were also tested for performance time, as online classification could not be performed because of software incompatibility: on a quad-core (Intel Core i7) 16-GB RAM personal computer, correlation-based feature extraction and classification required on average 0.18 seconds, of which 0.1 was due to the spectral estimate and peak selection and 0.05 to the CCA filtering process (time calculated over 3 electrodes, for a maximum of three pairs of data sets, as it grows linearly when increasing dimensionality); coherence extraction and classification for a single EEG recording required 0.77 seconds, of which 0.5 seconds were spent in the calculation of the AR coefficients (0.4 seconds) and the spectral estimates (0.1 seconds). As such we can conclude that both methods require an acceptable amount of time for online classification, which is negligible in the correlation-based classification and still reasonable for the coherence-based one. In particular, theoretical ITR can be computed according to the formula:

$$B = \log_2(N) + P \log_2(P) + (1 - P) * \log_2\left(\frac{1-P}{N-1}\right) \quad (5.3)$$

$$ITR = B * C$$

Where N is the number of possible classes, P is the probability the correct class is selected and C is the number of commands selected per minute. ITR (measured in bits/minute) is the gold standard for BCI studies and attests (for most recent SSVEP-based paradigms) values between 21.92 [51] and 27.15 bits/min [2]. These values however do not take into account the time the user requires to focus completely on the stimulus, nor the computation time for the acquisition and preprocessing of the EEG file. In this project, theoretical offline ITR can be calculated, considering the four-second long window of recording time and the running time of the classification code: average ITR for the correlation-based algorithm is 20.16 bits/min and 17.67 bits/min for the coherence approach.

If the acquisition and processing (around 0.60 s) are also taken into account, total recording and running time amounts to less than six seconds. Furthermore, subjects reported that the minimum time required for gaze shifting in free spelling trials was around 0.5 seconds. If a small overhead is added, $6.5 \div 7$ seconds is a reasonable estimate of time between the moment the user starts gazing at the blinking stimulus and the moment the result is displayed. As such, a value between 8.58 and 9.23 commands per minute can be spelled by the user, with real ITR for the correlation approach between 12.04 and 12.97 bits/min (maximum value for best subject: 25.88 bits/min) and for coherence between 10.2 and 10.97 bits/min (maximum value for best subject: 22.94 bits/min). These values are slightly below average, but could certainly be improved with different features combinations (see chapter 6 for discussion) and through code optimization, as the current implementation is not optimized for speed and could be exported in a different language with better performances.

Regarding the stability across trials, the subjects's alpha rhythm was similar in frequency and amplitude in both trials and roughly the same electrodes (or, more in general, the same cortical areas) had best SNR for the evoked potential. Furthermore, those who performed well at first try, kept a similar classification accuracy during the second trial, but did report being less annoyed by the flickers and were more confident in the execution of the task. On the other hand, subject who had low SSVEP response, still did not perform well the second time, probably because of the BCI illiteracy issue mentioned in chapter 2. Examples of this stability can be seen in Tables 5.10 and 5.11, which show the classifier performances for each stimulation frequency, using data from a single subject in two different trials (three recordings at same frequency per each trial).

Frequency	8	9	13	15	17	19	21
Accuracy (first trial)	1	0.67	1	1	0.67	0.33	0.33
Specificity (first trial)	0.89	1	0.94	0.89	0.94	1	1
Accuracy (second trial)	0.67	0.67	1	1	0.67	1	0.33
Specificity (second trial)	1	1	0.89	0.89	0.94	1	1

Table 5.10: Accuracy and specificity frequency by frequency across trials for the same subject in case of correlation-based classification. Note the similar misclassification rates for the higher frequencies.

Frequency	8	9	13	15	17	19	21
Accuracy (first trial)	0.33	0.67	1	1	1	1	0
Specificity (first trial)	1	0.94	1	0.78	0.94	1	1
Accuracy (second trial)	0.67	0.67	0.67	1	1	0.67	0.33
Specificity (second trial)	0.94	1	1	0.83	0.94	1	0.94

Table 5.11: Accuracy and specificity frequency by frequency across trials for the same subject in case of coherence-based classification. Note the low specificity at 15 Hz in both trials (i.e. the tendency of the classifier to incorrectly classify recordings as 15 Hz).

Lastly, the feedback from the test subjects was considered. After the experiment, several questions regarding the interface were asked, with an evaluation scale from 1 (poor/insufficient/unsatisfactory) to 5 (excellent/very satisfactory): all users gave the experience from average (3) to very positive ratings (5). Most of them did report some difficulties focusing on specific stimuli (usually low frequencies surrounding higher ones was found a bit distracting, as their blinking was noticeable even in peripheral vision), giving an average rates of 3 or 4 to the interface design. It is also worth noting that most subjects found it easier to concentrate toward the end of the experiment, because they got used to the interface and learned how to avoid being distracted by flashing stimuli in their peripheral vision. Subjects did not note any difficulties in free spelling tasks that required quick shifts of gaze; most reported the task as not particularly tiring (average grade: 4). Some subjects also noted that flickers set further apart would be helpful to avoid gaze shifting, but all praised the color choice: the orange and blue shades used were bright enough to contrast well with the black background but did not cause discomfort when gazing at them. These results could be used to improve the paradigm design and possibly increase performances by evoking higher quality SSVEP signals, i.e. closer to the best performing subject. This topic will be further explored and discussed in the final chapter 6.

Chapter 6

Conclusions

During the course of this project, a complete SSVEP-based BCI system was designed, implemented, tested and evaluated with different feature extraction techniques. First of all, the state of the art in paradigm and setup design was explored: SSVEP-based BCIs were chosen for their high ITR and asynchronous interface, which would be optimal for clinical use. Afterward, the designed paradigm was created using dedicated software, in order to ensure no timing issues would degrade the SNR of the evoked response. This resulted in a grid of alternating colored squares, with flickering frequencies of 8, 9, 13, 15, 17, 19 and 21 Hz. Since monitor refresh rate could be an issue for this purpose, an approximated frequency method was used for their implementation, so that an integer number of frames would be sufficient to simulate the frequency of interest (as found in recent SSVEP research literature). The experimental setup was then created according to the protocols in use in EEG recording, with an electrode cap mounted on the user, to record activity from the parieto-occipital (i.e. visual cortex) region. The 17 volunteers enrolled for this study were seated in front of a large monitor, focusing on a stimulus per request of the experimenter. Their EEG signals during a resting period and their evoked responses were then recorded. Before and after the experiment, a questionnaire was also administered to record demographics information and feedback on the experience. Subsequently, algorithms for the acquisition, preprocessing, feature extraction and classification of the EEG data were written in Matlab: first, a forward and backward filter was applied to remove components outside the band of interest. In particular, most of the power of the EEG spectrum is concentrated at lower frequencies even during rest recordings: this would polarize the classification results. As such, the bandpass filter was designed to maintain only a narrow frequency band and the presence of strong evoked potential harmonics (experimentally demonstrated) was exploited for the 8 Hz and 9 Hz recordings, since their fundamental component is cutoff from the filtered signal. The electrodes with highest SNR were subsequently selected using a form factor (ratio between the amplitude of the SSVEP and the unwanted frequency

components) on a recording of known frequency. Afterward, two classifiers were used: at first, a correlation-based classification was employed, using validated methodologies for CCA filtering and Pearson’s coefficient computation; then a new approach based on spectral coherence was implemented to solve phase delay issues encountered with the first one. This second method employed AR spectral estimation, which is a well developed technique for EEG recordings, encountered often in literature on the topic. Both approaches attempted to reduce computational requirements by using a classification in series, first using spectral based identification of the most likely candidate classes. This method used spectral identification of peaks close to the stimulation frequencies, as usually encountered in SSVEP identification software. However, the following steps (i.e. correlation-based or coherence-based classification) were designed to improve the accuracy of this classifier, so that ambiguous cases in which more than one frequency component had similar peaking could be placed in the correct target class. In order to do this, the information from the channels previously selected was reduced in dimensionality using CCA spatial filters, so that the resulting time series has maximal correlation with the training set and a combination of ideal sinusoids at the SSVEP frequency and their second harmonics. Two Pearson coefficients were then calculated between the test set of unknown SSVEP frequency and the training set and between the test set and ideal sinusoids of known frequency. An average correlation index was then computed and the highest one chosen as the representative of the target class. One weakness of this approach was found to be that correlation would naturally decrease if even small phase delay between the two sets was present. Consequently, spectral feature extraction was implemented. The coherence approach first undersampled the test set to reduce computing time of following processing, then computed the bivariate AR model between the one-channel test set and corresponding training set, with order determined using AIC. This model was then used to compute spectral estimates of the two time series and their cross-spectrum: from these measures, the coherence could be calculated. A peak in coherence was expected at frequencies in common between the two and the maximum normalized peak at the training set SSVEP frequency was selected as the one representative of the target class. The code was tested first on EEG simulations (designed as sums of zero-mean white noise and sinusoids, with similar SNR to SSVEPs in the frequency range of interest) and then on the real EEG data of the subjects. The experimental results were computed using confusion matrices for each subject, although seven candidates had to be discarded for either low recording quality or low evoked potentials. From them, mean indices of classification performances could be computed, first to optimize the number of channels and window length and then to compare the two methodologies. Four-second long recordings with, respectively, three and one channels were analyzed for ten subjects: the correlation-based classifier showed a median accuracy of $76.5 \pm 11.5\%$ (with Cohen’s κ index equal to 0.72) and the coherence-based one $71 \pm 25.5\%$ (κ equal to

0.43). Furthermore, a study on specificity and accuracy at each frequency demonstrated that the spectral classifier outperforms the temporal one at higher frequencies, although it has overall lower accuracy, while both have lower specificity for those frequencies that elicit the highest SSVEPs (13 and 15 Hz).

Considering the previous discussion, it can be said that the implemented system, on a software level, has results coherent with literature on the correlation-based classification, with excellent discrimination capabilities on medium SSVEP frequencies. On the other hand, regarding the new coherence-based approach (here implemented for the first time for SSVEP-based BCIs, to the best of the author's knowledge), classification results were positive, although slightly inferior than with the other methodology implemented. This approach only requires one channel, i.e. has less computational requirements, and it has been demonstrated to be more effective than correlation for frequencies higher than 15 Hz. Additionally, this approach is robust to time delays, making it well suited for an asynchronous interface where users can pace themselves as they prefer. Both these approaches do not require user-specific thresholding, so no training is required. Furthermore, they are useful in ambiguous classification cases (i.e. more than one frequency over threshold).

As such, a classifier using hybrid time-frequency feature vectors could be designed, so that their discrimination power is greater than that of the single features. In the future a study could be carried out to determine which method and category of classifier might be more appropriate for an online use of the system, since the offline analysis demonstrated the the designed software requires only short execution times, compatible with free-run spelling tasks. Another possibility to improve communication would be the implementation of an autocorrection system of the spelled word, so that classification errors for single letters inside the chosen word could be automatically corrected by the software (with a feedback to the users), as studies in literature have already suggested [52][53]. This would improve accuracy and could also be used for text prediction, so that users would have an option to select the complete target word after having spelled only a few letters, saving energy and time.

Regarding the BCI paradigm, ease of use and annoyance varied across the subjects: the interface could be improve using a bigger monitor, with increased stimulus size and distance between one flicker and its neighbors. Information gathered through the questionnaires during the test phase could be the first step toward the implementation of such an interface. In addition, by analyzing average accuracy for single frequencies, the paradigm can be further developed, avoiding higher frequencies (21 Hz, where the SNR showed an experimental steep decline, even in optimal subjects) and those too close to the alpha rhythm which could cause classification inaccuracy. A customization of the system could also be interesting, so that the most effective stimulation frequencies for the specific user can be used, considering the variation in evoked response encountered

across subjects.

An online test of the classifier is also needed, with communication between the interface controller and the acquisition software, for synchronization of the signals. This is fundamental to make the phase information of the EEG recordings available. In fact, SSVEP-based BCIs are limited by the finite number of frequencies that can be used for encoding, but a solution to this problem was proposed by Jia, C. et al in 2011 [46] with a mixed frequency and phase encoding. In this approach, after frequency is determined (for example using the classifiers described in this project), the phase is discriminated by studying the angle of the signal Discrete Fourier Transform on one channel, calculated at the frequency previously identified. Stimulation phases are usually 0 , $\pi/2$, π and $3/2 \pi$, but since there is a response latency between the visual stimulus and the SSVEP, there is a phase shift between the two (recently demonstrated to be constant across frequencies [50]). It is consequently necessary to create a training set where several trials at each frequency and phase are used to form clusters of phase responses, computed using an N-point Fast Fourier Transform (FFT) algorithm [54]. As such, this is reduced to a simple problem of classification in well-defined clusters and could be exploited to easily increase the number of symbols encoded in the paradigm without loss of accuracy, i.e. without using frequencies with low evoked response [46]. However, this method requires the definition of a specific protocol, to avoid any phase shift between the subject gazing at the stimulus and the recording window used for classification, which would compromise the clustering step.

It is worth noting that it was not possible to establish whether or not locked-in patients would find the implemented paradigm comfortable and whether it would allow them to spell each letter just shifting attention and not gaze, in case of absent voluntary eye movements. This was due to the fact that all test subjects were healthy and capable of independent eye movement. However, the stimulation interface was designed to be flexible and therefore could be rearranged to satisfy patients with severe disabilities or different needs and to allow a personalization of the system.

In conclusion, it can be said that the system implemented is a functional interface, acquisition and classification system that in the future could become a useful tool for BCI communication tasks: its flexible paradigm, through a new combined classification methodology that allows for classification of a wide range of frequencies, could be used for a higher number of targets and symbols, overcoming one of the greatest limitations of SSVEP systems for clinical applications.

Bibliography

- [1] G. Beverina, F. Palmas, “User adaptive bcis: Ssvep and p300 based interfaces,” *PsychNology Journal*, vol. 1, pp. 331–354, Dec 2003.
- [2] A. Amiri, S. Rabbi, *Brain-Computer Interface Systems - Recent Progress and Future Prospects*, ch. 10: A Review of P300, SSVEP, and Hybrid P300/SSVEP Brain-Computer Interface Systems, pp. 195–214. Intech.
- [3] H. Muhl, C. Gurkok, “Bacteria hunt - evaluating multi-paradigm bci interaction,” *J Multimodal User Interfaces*, vol. 4, Aug 2010.
- [4] Y. Wang, Y. Wang, “Visual stimulus design for high-rate ssvep bci,” *Electronics Letters*, vol. 46, pp. 1057–1059, Jul 2010.
- [5] R. Wang, Y. Wang, “Brain-computer interface based on the high-frequency steady-state visual evoked potential,” in *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*, May 2005.
- [6] M. Nakanishi, “A high-speed brain speller using steady-state visual evoked potentials,” *Int. J. Neur. Syst.*, vol. 24, p. 1450019, Jun 2014.
- [7] M. Formisano, R. D’Ippolito, “Vegetative state, minimally conscious state, akinetic mutism and parkinsonism as a continuum of recovery from disorders of consciousness: an exploratory and preliminary study,” *Funct Neurol.*, vol. 26, pp. 15–24, Jan-Mar 2011.
- [8] R. Ficke, *Digest of Data on Persons with Disabilities, 1991*. Washington, DC: U.S. Department of Education, National Institute on Disability and Rehabilitation Research, 1991.
- [9] S. Schnakers, C. Majerus, “Cognitive function in the locked-in syndrome,” *J Neurol.*, vol. 255, pp. 323–330, Mar 2008.
- [10] K. Majaranta, P. Raiha, “Twenty years of eye typing: Systems and design issues,” in *Proceedings of Eye Tracking Research and Applications*.

- [11] R. Leow, “Development of a steady state visual evoked potential (ssvep)-based brain computer interface (bci) system,” in *International Conference on Intelligent and Advanced Systems*.
- [12] C. Ang, KK Guan, “Brain-computer interface in stroke rehabilitation,” *J Comput. Sci. Eng.*, vol. 7, pp. 139–146, Jun 2013.
- [13] N. Birbaumer, “Communicating with a locked-in patient,” May 2011.
- [14] N. Wolpaw, JR Birbaumer, “Brain–computer interfaces for communication and control,” *Clin. Neurophysiol.*, vol. 113, Mar 2002.
- [15] A. Morshed, BI Khan, “A brief review of brain signal monitoring technologies for bci applications: Challenges and prospects,” *J Bioeng Biomed Sci*, vol. 4, p. 1000128, May 2014.
- [16] M. Lotte, F. Congedo, “A review of classification algorithms for eeg-based brain–computer interfaces,” *J Neural Eng*, vol. 4, pp. R1–R13, Mar 2007.
- [17] F. Pfurtscheller, G. Lopes Da Silva, “Event-related eeg/meg synchronization and desynchronization: basic principles,” *Clin. Neurophysiol.*, 1999.
- [18] J. Allison, B. Faller, *Brain-Computer Interfaces: Principles and Practice*, ch. BCIs that use steady state visual evoked potentials or slow cortical potentials.
- [19] N. Birbaumer, N. Ghanayim, “A spelling device for the paralysed,” *Nature*, vol. 398, Mar 1999.
- [20] C. Lin, Z. Zhang, “Frequency recognition based on canonical correlation analysis for ssvep-based bcis,” *IEEE Trans. Biomed. Eng.*, vol. 53, pp. 2610–2614, Dec 2006.
- [21] X. Bin, G. Gao, “An online multi-channel ssvep-based brain-computer interface using a canonical correlation analysis method,” *J Neural. Eng.*, vol. 6, Jun 2009.
- [22] E. Farwell, LA Donchin, “Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials,” *Electroencephalogr. Clin. Neurophysiol.*, 1988.
- [23] J. Jin, “Optimized stimulus presentation patterns for an event-related potential eegbased brain-computer interface,” *Med. Biol. Eng. Comput.*, 2011.
- [24] A. Brunner, P. Ritaccio, “Rapid communication with a “p300” matrix speller using electrocorticographic signals (ecog),” *Front. Neurosci.*, 2011.
- [25] Z. Chen, X. Chen, “A high-itr ssvep based bci speller,” *Brain-Comp. Interfaces*, vol. 1, pp. 181–191, Sep 2014.

- [26] W. Spuler, M. Rosenstiel, “Online adaptation of a c-vep brain-computer interface (bci) based on error-related potentials and unsupervised learning,” *PLoS One*, 2012.
- [27] G. Calhoun, “Eeg-based control for human-computer interaction,” in *Proc. 3rd Annu. Symp. Human Interaction with Complex Systems*.
- [28] M. Middendorf, “Brain-computer interfaces based on the steady-state visual-evoked response,” *IEEE Trans. Rehab. Eng.*, vol. 08, pp. 211–214, Jun 2000.
- [29] B. Guger, C. Allison, “How many people could use an ssvep bci?,” *Front. Neurosci.*, vol. 06, Nov 2012.
- [30] T. Thiemann, “An investigation of the test process used to date for determining the response time of an lcd monitor, known as input lag.”
- [31] D. Krusienski, DJ McFarland, “An evaluation of autoregressive spectral estimation model order for brain-computer interface applications,” in *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*, Aug 2006.
- [32] R. Nunez, PL Srinivasan, “Eeg coherency: I: statistics, reference electrode, volume conduction, laplacians, cortical imaging, and interpretation at multiple scales,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, pp. 499 – 515, Nov 1997.
- [33] M. Koles, ZJ Lazaret, “Spatial patterns underlying population differences in the background eeg,” *Brain topography*, 1990.
- [34] G. Hakvoort and B. Reuderink, “Comparison of psda and cca detection methods in a ssvep-based bci-system,” Feb 2011.
- [35] G. Zhang, Y. Zhou, “Frequency recognition in ssvep-based bci using multiset canonical correlation analysis,” *Int J Neural Sys.*, vol. 24, no. 04, p. 1450013, 2014.
- [36] T. Allison, B. Luth, “Bci demographics: How many (and what kinds of) people can use an ssvep bci?,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, pp. 107–116, Apr 2010.
- [37] H. Martinez, P. Bakardjian, “Fully online multicommand brain-computer interface with visual neurofeedback using ssvep paradigm,” *Comput. Intell. Neurosci.*, 2007.
- [38] C. Allison, B. Neuper, *Brain-Computer Interfaces: Principles and Practice*, ch. Could anyone use a BCI?, pp. 35–54. Ed. Springer.
- [39] A. Kubler, “Brain-computer interfaces and communication in paralysis: extinction of goal directed thinking in completely paralysed patients?,” *Clin. Neurophysiol.*, vol. 119, pp. 2658–2666, Nov 2008.

- [40] J. S. R. van Gerven, M. Farguhar, “The brain-computer interface cycle,” *J Neural Eng.*, vol. 6, no. 04.
- [41] A. Fatourech, M. Bashashati, “Emg and eog artifacts in brain computer interface systems: A survey,” *Clin. Neurophysiol.*, 2007.
- [42] A. Xing, S. McDaid, “Impact of electrode positions and harmonic frequency components in ssvep-based bcis,” *IJISAE*, 2015.
- [43] L. Bianchi, L. Quitadamo, “Performances evaluation and optimization of brain computer interface systems in a copy spelling task,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, pp. 207–216, Jun 2007.
- [44] I. Neurobehavioral Systems, “Presentation: Stimuli: Picture timing accuracy.”
- [45] M. Hansson-Sandsten, “Evaluation of the optimal lengths and number of multiple windows for spectrogram estimation of ssvep,” *Med. Eng. Phys.*, vol. 32, pp. 372 – 383, May 2010.
- [46] C. Jia, “Frequency and phase mixed coding in ssvep-based brain-computer interface,” *IEEE Trans. Biomed. Eng.*, 2011.
- [47] K. C. T. Fabri, SG Camilleri, *Biomedical Engineering, Trends in Electronics, Communications and Software*, ch. Parametric Modelling of EEG Data for the Identification of Mental Tasks, pp. 367–386. Intech.
- [48] A. Baselli, G. Porta, “Spectral decomposition in multichannel recordings based on multivariate parametric identification,” *IEEE Trans. Biomed. Eng.*, vol. 44, pp. 1092–1101, Nov 1997.
- [49] J. Bassani, T. Nievola, *Brain Inspired Cognitive Systems 2008*, vol. 657, ch. Brain-Computer Interface Using Wavelet Transformation and Naïve Bayes Classifier, pp. 147–165. Ed. Springer.
- [50] M. Nakanishi, “Generating visual flickers for eliciting robust steady-state visual evoked potentials at flexible frequencies using monitor refresh rate,” *PLoS ONE*, vol. 9, Sep 2014.
- [51] P. V. I. Gembler, F. Stawicki, “Spectral decomposition in multichannel recordings based on multivariate parametric identification,” *Front Neurosci.*, vol. 9, Dec 2015.
- [52] A. Mora-Cortes, “Language model applications to spelling with brain-computer interfaces,” *Sensors*, vol. 14, pp. 5967–5993, Mar 2014.

- [53] T. Dalbis, “A predictive speller controlled by a brain-computer interface based on motor imagery,” *ACM Trans. on Computer-Human Interaction*, vol. 19, pp. 20–25, Oct 2012.
- [54] W. Klonowski, “Everything you wanted to ask about eeg but were afraid to get the right answer,” *Nonlinear Biomedical Physics*, vol. 3, no. 1, pp. 1–5, 2009.
- [55] D. Hardoon, “Canonical correlation analysis; an overview with application to learning methods,” *Department of Computer Science Royal Holloway, University of London*, 2003.

Appendix A

Canonical Correlation Analysis

An exhaustive definition of CCA was given by D. R. Hardoon in the technical report *Canonical correlation analysis: An overview with application to learning methods* (2003) [55].

Proposed by H. Hotelling in 1936, canonical correlation analysis can be seen as the problem of finding basis vectors for two sets of variables such that the correlation between the projections of the variables onto these basis vectors are mutually maximised. Correlation analysis is dependent on the co-ordinate system in which the variables are described, so even if there is a very strong linear relationship between two sets of multidimensional variables, depending on the co-ordinate system used, this relationship might not be visible as a correlation. Canonical correlation analysis seeks a pair of linear transformations one for each of the sets of variables such that when the set of variables are transformed the corresponding co-ordinates are maximally correlated.

In particular, CCA finds the weight vectors W_x and W_y , called canonical coefficient vectors, that maximize the correlation between the two data sets, i.e. the cross-correlation coefficient ρ :

$$\max_{W_x, W_y} [\rho(x, y)] = \frac{E[W_x^T X Y^T W_y]}{\sqrt{E[W_x^T X X^T W_x] E[W_y^T Y Y^T W_y]}}$$

and repeats the operation N times to find new basis vectors independent from the ones just calculated, where N is the minimum rank between X and Y , the data matrices.

Acknowledgements

I would like to express my gratitude to my thesis advisor, prof. Anna Maria Bianchi, and my tutor, Dr. Giulia Tacchino, for their invaluable help.

My heartfelt thanks go to the Engineering Department of the University of Aarhus (Denmark) who collaborated on this project from its inception and in particular to the professor Preben Kidmose who assisted in its development.

I also want to thank all the participants in the study, for their willingness to take part in the experimentation and my friends for their help and encouragement.

Lastly, I would like to thank my family, for their unwavering support all these years.