ECOC BASED MULTI-CLASS CLASSIFICATION IN BRAIN COMPUTER INTERFACES WITH SSVEP

by
SANDRA SAGHIR

Submitted to the Graduate School of Social Sciences
in partial fulfilment of
the requirements for the degree of Master of Electronics Engineering

Sabancı University
December 2020

ECOC BASED MULTI-CLASS CLASSIFICATION IN BRAIN COMPUTER INTERFACES WITH SSVEP

APPROVED BY:

Date of Approval: 15 December 2020

ECOC BASED MULTI-CLASS CLASSIFICATION IN BRAIN COMPUTER INTERFACES WITH SSVEP

SANDRA SAGHIR

Electronics Engineering, MSc. Thesis, December 2021

Thesis Supervisor: Assist. Prof. Dr. Hüseyin Özkan

Co-advisor: Assist. Prof. Dr. Nihan Alp

Keywords: Steady state visually evoked potentials (SSVEP), brain-computer interfaces (BCI), electroencephalography (EEG), error correcting output codes (ECOC), multi-class classification

Abstract

Brain-Computer Interfaces (BCIs) based on steady-state visual evoked potential (SSVEP) responses are among the most frequently used non-invasive BCI systems due to their feasibility, portability, and low cost. SSVEPs are the brain responses to flickering visual stimuli at a specific frequency. One of SSVEP's critical applications is SSVEP-based BCI speller; this system allows disabled people to communicate directly by using their brain signals without dependence on speech production. An SSVEP-based BCI speller incorporates a variety of flickering characters or numbers. Therefore, decoding brain activities for an SSVEP-based BCI speller requires solving a multi-class classification problem. Over the last few years, various studies have attempted to achieve higher frequency recognition accuracy and faster information transfer rates to enhance the recognition performance. This thesis employs an ensemble method called Error-Correcting Output Codes (ECOC) to tackle the above-mentioned multi-class classification problem. To the best of our knowledge, the ECOC framework has not been explored for the SSVEP classification problems to date. We present an extensive set of comparisons among four prominent ECOC coding matrix designs, one-vs-all (OVA), one-vs-one (OVO), random dense, and random sparse. Furthermore, three feature extraction methods are investigated to evaluate the overall performance of such designs. The utilized feature extraction methods

include Canonical Correlation Analysis (CCA), Power Spectrum Density Analysis (PSDA) via Welch's method, and Correlated Components Analysis (CORRCA). Using the ECOC ensemble method improves the general performance compared to the standard methods such as standard CCA and standard CORRCA. Moreover, the results indicate the superiority of the feature extraction method CORRCA especially for a short time window and the OVA coding matrix design. In conclusion, The presented approach has the ability to incorporate high-performance BCI speller systems based on SSVEP.

DHGUP İLE BEYİN BİLGİSAYAR ARAYÜZLERİNDE HDÇK TABANLI ÇOK SINIF SINIFLANDIRMA

SANDRA SAGHIR

Elektronik Mühendisliği, Master Tezi, 2021

Tez Danışmanı: Assist. Prof. Dr. Hüseyin Özkan

Anahtar Kelimeler: Durağan hal görsel uyarılmış potansiyel (DHGUP), beyin-bilgisayar arayüzü (BBA), elektroensefalografi (EEG), Hataya Dayanıklı Çıktı Kodları (HDÇK), çok sınıflı sınıflandırma

ÖZET

salınım yapan görsel uyaranlara tepkisidir. DGHUP' ın önemli uygulamalarından birisi de DHGUP bazlı BBA heceleyicidir. Bu sistem engelli bireylerin çevreleriyle, konuşmadan Durağan hal görsel uyarılmış potansiyel (DHGUP) bazlı beyin-bilgisayar arayüzü (BBA), makul, taşınabilir ve daha az masraflı olması sebebiyle girişimsel olmayan (non-invaziv) BBA sistemleri arasında kullanılan en yaygın yöntemdir. DHGUP sinyalleri, beynin belirli frekansta , sadece beyin sinyallerini kullanarak iletişim kurabilmelerine olanak sağlar. DHGUP bazlı BBA heceleyicide birçok salınım yapan karakter ve sayılar kullanır. Bu yüzden, beyin aktivitelerini anlamlandırmak, çok sınıflı bir sınıflandırma problemidir. Geçtiğimiz yıllarda, birçok çalışma daha iyi bir performans elde edebilmek için, yüksek doğruluk oranlı frekans kestirimi ve yüksek bilgi aktarım hızı elde etmeye çalışmışlardır. Bu tezde, bahsedilen çok sınıflı sınıflandırma problemini çözmek için bir torbalama yöntemi olan, Hataya Dayanıklı Çıktı Kodları (HDÇK) kullanılmıştır. Bildiğimiz ka.darıyla bugüne kadar, HDÇK yöntemi DHGUP sinyallerini sınıflandırma probleminde kullanılmamıştır. HDÇK matris tasarım algoritmalarından 4 tanesinin, bire-

bir kodlama, bire-hepsi kodlama, yoğun ve ayrık rastgele kodlama, arasında kapsamlı bir karşılaştırma sunulmuştur. Ayrıca üç tane öznitelik çıkarma metodu, bu matris tasarımlarının performanslarını incelemek için kullanılmıştır. Kullanılan öznitelik çıkarma metotları, kanonik korelasyon analizi (KKA), korelasyonlu parçacık analizi (KORRPA), Welch metodu ile güç spektral yoğunluğu analizidir (GSYA). HDÇK metodu, standart metotlara göre, örneğin standart KKA ve standart KORRPA metotlarına göre, performansı iyileştirmiştir Sonuçlar, KORRPA öznitelik çıkarma metodunun, özellikle kısa zaman aralıklarında, diğer öznitelik çıkarma metotlarına göre, bire-hepsi matris tasarımının diğer matris tasarımlarına göre, daha iyi performans verdiğini göstermiştir. Sonuç itibariyle, sunulan yöntem, yüksek performanslı DHGUP bazlı BBA heceleyici sistemlerinin uygulanmasında, bir potansiyele sahiptir.

ACKNOWLEDGEMENTS

I would like to take this opportunity to convey to my thesis advisor Assist. Prof. Dr. Hüseyin Özkan my profound and sincere appreciation for giving me the opportunity to work with him and providing me with exceptional supervision, invaluable guidance, and endless support. He has taught me how to carry out the research framework and with his support and patience, which has enabled me to complete my thesis journey.

I would also like to thank my thesis co-supervisor Assist. Prof. Dr. Nihan Alp for her endless help, support, gaudiness and valuable suggestions during each meeting.

I would like to thank my jury members Prof. Özgür Gürbüz, Prof. Berrin Yanıkoğlu and Assoc. Prof. Şuayb Ş. Arslan to join my thesis presentation and for valuable discussion.

I also would like to thank Osman Berke Guney, Begum Sonmez and Serkan Musellim for providing their valuable insights to my work.

I would like to thank my parents for their unconditional love, efforts and support throughout in shaping up my life. I would like to thanks also my sisters and brother for their encouragement and support.

I am also thankful for my lab-mates, and my friends who join me in my journey Mastaneh Torkamani Azar, Sara Atito Ali Ahmed, Aysa Jafari and Naida Fetic.

I would like to acknowledge the financial support provided by Sabanci university Dean office for granting me the scholarship and opportunity for my M.Sc. studies. I am grateful for this scholarship and hoping that the program can keep on to support many students to accomplish their dreams.

All the glory and praise to God, for bestowing upon me strength, endurance and inspiration.

Dedication
To my beloved parents

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BCI: Brain-Computer Interface

SSVEP: Steady-State Visual Evoked Potential

EEG: Electroencephalogram

ITR: Information Transfer Rate

SVM: Support Vector Machine

ECOC: Error-Correcting Output Codes

CCA: Canonical Correlation Analysis

CORRCA: Correlated Components Analysis

PSDA: Power Spectrum Density Analysis

1. Introduction to Thesis Topic

A direct contact path between the human brain and an external system is presented through brain-computer interfaces (BCIs). Thus, the BCI system translates the brain signals into commands such that a device like a computer executes. By using the brain signals without the need for muscle movement, this system helps disabled people to regulate and communicate with their environments.

One of the recent BCI paradigms is based on steady-state visual evoked potential (SSVEPs) in which a user is placed in front of a computer screen, and this screen displays several flickering targets at various frequencies. The essence of this method is when a user gazes at a particular stimulus that flickers at a specific frequency, the EEG signals that are recorded from the scalp generates electrical activity at the same frequency and its harmonics.

In other words, SSVEPs are the brain's responses to the repetitively flickering visual stimuli that flash at various frequencies. Furthermore, it is a photomotor response characterized by sinusoidal-like waveforms at the frequency of the flickering stimulus and its multiple frequencies (harmonics) [1].

SSVEP-based BCI systems have gained interest over the last several years due to several advantages: high information transfer rate (ITR) and little user training. SSVEP spellers are one of the most widely used SSVEP-based BCI systems [2]. These systems provide a possible way of communication for the people who suffer from motor neuron disease (MND) or amyotrophic lateral sclerosis (ALS). Therefore, researchers and developers seek to enhance this system performance to build an efficient and high-speed SSVEP-based BCI speller.

1.1 Scope and Motivation

In this thesis, we are motivated to improve the SSVEP-based BCI speller performance. In other words, we aim to build a high-speed BCI speller with high-frequency recognition accuracy. In fact, several factors can affect SSVEP speller systems' performance, such as the number of flickering stimuli, the number of channels, electrode locations, and signal processing methods for target identification. In this study, we focus on investigating a novel approach for SSVEP target identification. Therefore, an ECOC framework is applied to deal with multi-class classification. Thus, the user can efficiently elect a specific target from several possibilities. We select the most convenient and efficient feature extraction methods for SSVEP-based BCI in the literature to evaluate the ensemble ECOC method. Three different feature extraction methods are included: canonical correlation analysis (CCA), power spectrum density analysis (PSDA) via Welch's method, and correlated components analysis (CORRCA). The ECOC method's performance with different feature sets is measured by computing the classification accuracy averaged across all subjects based on a publicly available large SSVEP speller benchmark dataset, and this dataset is recorded from 35 subjects. Moreover, another measurement is used to evaluate the performance, which is the information transfer rate (ITR), and this measurement determines the amount of transformed information and the speed of the SSVEP speller system.

1.2 Thesis Contributions

This thesis's contributions can be summarized as follows:

• This study demonstrates the applicability of merging novel and state-of-the-art techniques with a complicated task of dynamic brain decoding. It researches and improves the way for building reliable and portable brain speller systems that can enhance the quality of life for patients suffering from various neuromuscular issues.

• In this study, an ensemble method called Error-Correcting Output Codes (ECOC) is investigated to solve the multi-class classification problem. To the best of our knowledge, the ECOC paradigm has not been applied previously to the SSVEP classification problems.

• To analyze the performance of SSVEP-based BCI speller with the ECOC ensemble method, we use three different feature extraction methods, which are canonical correlation analysis (CCA) and power spectrum density analysis (PSDA) via Welch's and correlated components analysis (CORRCA).

- An extensive set of comparisons is performed among the most widely known ECOC coding matrix designs, one-vs-all (OVA), one-vs-one (OVO), random dense, and random sparse. The overall performance is measured in terms of classification accuracy and information transfer rate (ITR).

- As a result, the ECOC framework improves SSVEP-based BCI speller's performance compared to the standard methods like standard CCA and standard CORRCA. Furthermore, we compare several ECOC coding matrix designs, OVA, OVO, random dense, and random sparse, and the results show that the coding matrix designs, OVA and random sparse, have superior performance than others. Moreover, for each feature extraction method using the ECOC framework, we report the performance of coding matrix designs with various data lengths from 0.5 s to 5s. Consequently, using the OVA coding matrix design with CORRCA features leads to more reliable results compared to other alternatives.

1.3 Thesis Outline and Organization

The remainder of the thesis is organized as follows:
Chapter 2 provides general background on brain-computer interfaces (BCIs), EEG signals, steady-state visually-evoked potential (SSVEP). Furthermore, it gives an introduction of SSVEP-BCI speller and also states some related work.
Chapter 3 provides the problem description and the general contribution of this thesis.
Chapter 4 explains the general framework of feature extraction methods, Canonical Correlation Analysis (CCA), Power Spectrum Density Analysis (PSDA) via Welch's method, and Correlated Components Analysis (CORRCA).
Chapter 5 gives a background on the ECOC method and discusses the most well known ECOC matrix designs, OVA, OVO, random dense, and random sparse.
Chapter 6 provides an analysis of ECOC structures with the three feature extraction methods CCA, PSD, and CORRCA. Furthermore, this chapter presents the final results in terms of classification accuracy and the information transfer rate.
Finally, the thesis is concluded in **Chapter 7**.

2. Background on BCIs and SSVEP-based BCI Spellers

This chapter provides the basic concepts of brain-computer interfaces (BCIs), EEG signals, steady-state visually-evoked potential (SSVEP), and SSVEP-based BCI spellers. Moreover, it includes a general review of some related works and previous methods.

2.1 Brain-Computer Interfaces (BCIs)

BCI technology was introduced at the beginning of the 1970s [3]. Nowadays, many research areas have focused on enhancing the quality of life for individuals who suffer from stroke, Parkinson's disease, and Amyotrophic Lateral Sclerosis (ALS) by allowing them to gain some control in order to contact their external environment [4]. The BCI system has been developed to permit an alternative communication method for disabled people by interpreting their brain activity.

Thus, brain-computer interface (BCI) is a process that grants direct interaction between the brain and the external world like computers or any device. In this process, the brain activities are recorded and then translated into commands without using any muscular activity [5].

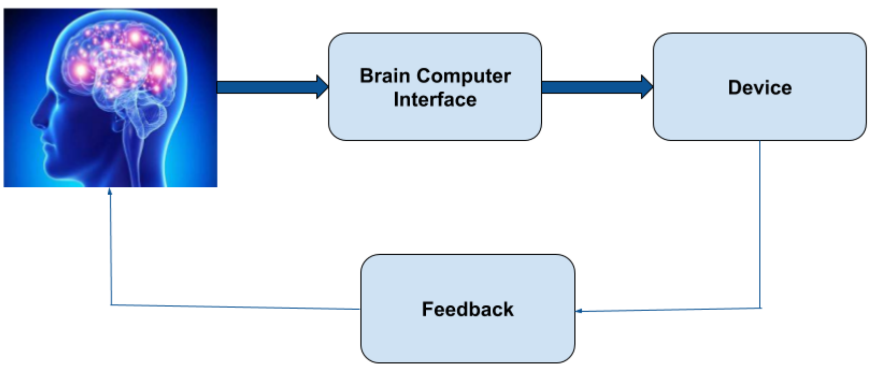

Figure 2.1 A typical BCI framework

In general, a typical BCI system has several consecutive stages. The first stage is signal acquisition. For signal acquisition, two main methods are primarily utilized: invasive and non-invasive BCI signal acquisition. Most non-invasive BCI technology mainly uses Electroencephalography (EEG) to record brain activities [6]. In EEG, the signals are compared between two electrodes as a voltage transition, and the electrodes are positioned over the human scalp at various locations. After collecting the signals, some preprocessing methods are usually applied to clean the contaminated signals that are resulted from noise and artifacts. The relevant features are then extracted and decoded into commands that the device can understand, and this can be achieved by using an efficient classification method. Some systems can provide feedback, as shown in Fig. 2.1. The feedback is commonly presented in a visual or acoustic version. For example, it can give a beep sound after encoding each command. In addition, some applications use the feedback to keep the participant concentrated and focused during the experiment.
A conventional BCI process consists of three main steps: signal preprocessing, feature extraction, and classification; each step is crucial to obtain a feasible BCI system.

- **Signal Preprocessing:**

The raw data collected from the scalp of the brain is often a signal contaminated by noise due to eye movements, muscle activities, and external resources. The preprocessing step is an essential process to eliminate the unwanted data and artifacts from the signals [7]. These artifacts may be generated due to physiological or non-physiological sources. The physiological sources can be eye and muscle activities, and non-physiological causes such as impedance mismatch, power-line coupling, etc. Different kinds of filters are usually applied for preprocessing the signals, like spatial and spectral filters, to compress the generated contaminated signals. As a consequence, many studies have shown that applying the prepossessing method to raw data could significantly affect the BCIs system's performance, and especially this can have a numerous influence on the classification accuracy [8, 9].

- **Feature Extraction:**

Feature extraction is a procedure of dimensionality reduction by getting manageable information from the raw data. In other words, the obtained feature set resulted from the feature extraction method is a reduced set of features that compiles the most valuable information from the initial set of features [10]. The feature extraction process is usually applied to the signals after some preprocessing methods, and based on BCI applications or area of interest, the appropriate feature extraction method is selected.

• **Classification:**
The classification method is commonly chosen based on the obtained feature set, and it bases on matching the features coming from signals to their corresponding commands. One of the crucial measurements to evaluate the BCI system performance is the classification accuracy, and it is calculated by The number of correctly categorized commands compared to the total number of system-categorized commands. Various frequency recognition methods related to SSVEP-based BCI speller of some previous studies will be discussed and compared in Section 2.3.3.

2.2 Types of BCI Systems

There are two kinds of BCI systems for signal acquisition process:

• **Invasive BCIs:**
In invasive BCI, the signals are recorded directly from the cortex; specifically, electrodes are implemented into the grey matter during neurosurgery. High-quality signals are extracted in this system due to the direct connection with the brain [6]. However, it has some disadvantages, like it is prone to scar-tissue build-up, or the patient's body may not accept the implemented object [11]. Besides, it is costly and hard to implement, such as electrocorticography (ECoG) [12]. Therefore, most of the experiments with invasive BCI is often conducted for some medical purposes.

• **Non-invasive BCIs:**
In the non-invasive BCI, sensor electrodes are usually used for the signals acquisition, where these sensors are placed on the scalp of the brain. The signals in non-invasive BCI systems have a lower signal to noise ratio compared to the invasive system [13]. However, it is more feasible, practical, and easy to implement. Fig. 2.2 shows different types of BCI system with their electrodes placement. In an invasive system, electrodes are implemented in the cortical surface of the patient's brain. In contrast, the electrodes are located on the brain surface in a non-invasive system. The most common BCI systems are generally used the sensor electrodes to record the brain activity like magnetoencephalography (MEG)) [14], functional magnetic resonance imaging (fMRI) [15], functional near-infrared spectroscopy (fNIRS) [16], positron emission tomography (PET) [17], and electroencephalogram (EEG), which is considered one of the most popular non-invasive technique [18].

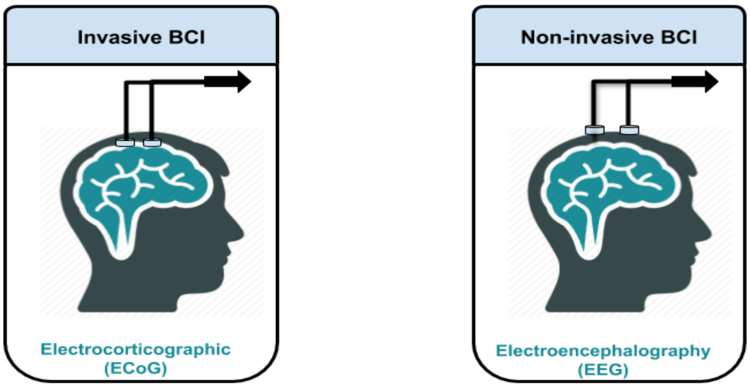

Figure 2.2 A comparison between invasive and non-invasive BCI

2.2.0.1 Electroencephalogram (EEG)

The electroencephalogram (EEG) is utilized as the foundation for BCI [19]. Additionally, EEG is one of the most preferred non-invasive signal acquisition tools due to its various advantages. Those advantages include simplicity in usage and implementation. Besides, using EEG for signal acquisition doesn't expose the patient's body to any magnetic field or x-ray. Therefore, it doesn't contain any side effects, and also the experiments can be readily conducted several times. Furthermore, it has a low cost compared to other non-invasive tools. EEG measures the electrical activity of the brain using electrodes located on the scalp of the brain. Thus, it provides a high temporal resolution. Moreover, it can be helpful to easily evaluate how brain function can change in response to stimuli, and can also be advantageous in measuring irregular brain activity, as in epileptic seizures [20]. Fig. 2.3 illustrates a brief example about EEG signals. In an EEG-based BCI framework, the recorded signals from an EEG amplifier are first preprocessed and then classified to decode the intent of the user. Therefore, EEG signals can present the input signals for several applications like robotic arm control [21] and cursor control [22].

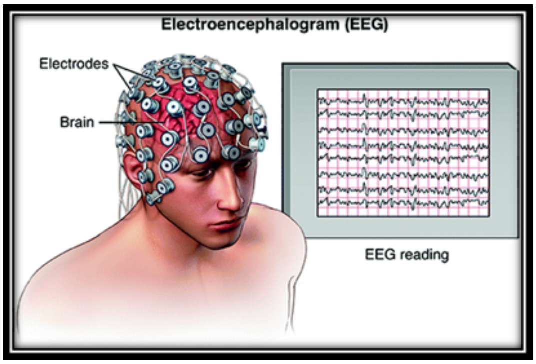

Figure 2.3 An example of EEG signals, taken from [23]

2.3 Steady State Visually-Evoked Potential (SSVEP)

Steady-state visual evoked potentials (SSVEP) are the brain responses in electroencephalographic (EEG) signals to repetitive flickering visual stimuli. In the SSVEP experiment, the subject is placed at a specific distance from a screen that displays stimuli. These stimuli are flickering at various frequencies, and each frequency corresponds to an individual command. Moreover, by gazing at a particular stimulus, the user can elect the requisite command. The frequency of responses corresponds with the frequency of stimulation, harmonics, and subharmonics for the corresponding stimulus of the subject.

BCI has a plethora of paradigms, including motor imaginary, P300, and SSVEP. However, over the last few years, SSVEP-based BCI drew attention due to the several benefits such as high information transfer ratio (ITR), ease of system configuration, and little time for training [24].

The general process of the SSVEP-based BCI system is illustrated in Fig. 2.4.

After collecting the user's data using EEG for a non-invasive BCI system, the signals are preprocessed to reduce the external artifacts, and then the essential information is extracted from the signals. The next phase is to choose an appropriate classification method in order to relate the extracted features to its corresponding class, and the last step is translating the final results to be understood by a device.

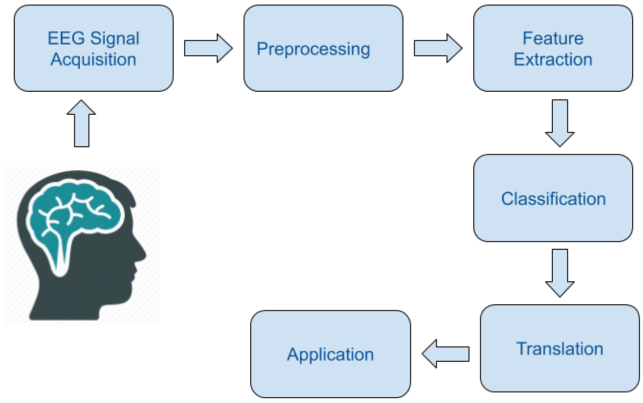

Figure 2.4 A general paradigm for SSVEP-based BCI process

• **SSVEP-based BCI Speller**

One of the primary applications for SSVEP-based BCI fields is the SSVEP-based BCI speller. Fig. 2.5 gives a brief example of an SSVEP-based BCI experiment. The subject is placed in front of a screen monitor that presents certain characters and symbols. Each command is flickering in a specific frequency, and the subject is gazing into certain stimuli that flicker in a particular frequency.

Figure 2.5 An example of SSVEP-based BCI speller experiment, taken from [25].

In the last few years, researchers and developers have investigated the SSVEP-based BCI in several aspects to enhance the SSVEP paradigm.
Some principal characters have significant effects on SSVEP-based BCI, such as
(a) The effect of the target stimuli design (stimulus display and number of targets),
(b) The number of channels and electrodes locations, and
(c) Signal processing methods for target identification.
Other than these aspects, it is also stated that the participants' physical conditions

might affect SSVEP performance, such as their age [26].

2.3.1 The Effect of the Target Stimuli Design

One of the critical components that can affect SSVEP-based BCI performance is designing the appropriate stimuli for the BCI system. The number of stimuli and its location have a significant impact on the SSVEP-based BCI. Generally, the visual stimulator is flickering targets that it is possible to present using flashing light-emitting diodes (LEDs) or displayed on a liquid crystal display (LCD) like a monitor of device or computer [1]. Using a computer monitor is convenient, especially for programming feedback. However, a wide range of stimulation frequencies and, thus, a much broader frequency band are needed for the vast number of targets. Nevertheless, humans exhibit superior SSVEP responses only in a specific range of frequencies [27]. The computer monitor utilizes as a stimulator that is only able to generate a confined range of frequencies due to the limitation of the refresh rate of the screen. Thus, the available stimulation frequencies for SSVEP-based BCI was restricted [28]. Therefore, several approaches have been used to resolve this limitation. One of the proposed methods is called Multiple Frequencies Sequential Coding (MSFC), in which multiple frequencies are used sequentially to code the targets instead of only a constant frequency [29]. The MFSC method is based on permutation theory. For example, if N frequencies are used for target coding and M is the length of the coding sequence, then N^M permutation sequences can be coded in this method, unlike the single frequency coding method that can code only N targets.

Moreover, one study introduced the combination of phase and frequency to design the stimuli [30]. Hence, a phase shift was added to target encoding to improve the number of stimuli with a limited range of frequency. The design of the target stimulus was as follows; 6 of the stimulus are flickering at frequency of 10 Hz with a phase of 60° degree between neighbouring stimuli, 5 other stimulus are flickering with a frequency of 12 Hz with phase 72° between neighbouring stimuli, and the last 4 stimulus are flickering with 15 Hz frequency with phase of 90° between nearby stimuli. The results indicate that using mixed-phase and frequency coding improved the ITR rate.

An alternative method is named Frequency Shift Keying (FSK)-Modulated Visual Stimuli [31], in which a codeword represents the visual stimuli. In other words, the frequencies of the flickering stimulus are identified by binary digits. There are two main parts of the FSK technique: encoding and modulation. This technique's output target determination also consists of two steps: demodulation and decoding.

This technique can overcome the restriction of the number of stimuli; as the length of the codeword increases, more commands can be generated. The outcome showed that eight subjects out of 10 subjects obtained adequate precision using FSK modulation.

On the other hand, the performance of SSVEP-based BCI can be influenced by stimulus properties such as the color, scale, and location of visual stimuli [24]. Some studies examined the effect of the luminance and chromatic properties of the target flickering on the performance of BCIs. In one study, nine flickering targets were tested with three different conditions for optimizing the stimulation design: luminance, color, and (luminance and color) conditions for each target [28]. As a result, the combination of chromatic and luminance has enhanced the classification accuracy for SSVEP-based BCI.

Furthermore, the refresh rate of the monitor can also affect the classification performance of the SSVEP-based BCI. One of the studies compares the performances of SSVEP spellers when two different values (120Hz - 75Hz) are used as the refresh rate of the computer [1]. The outcome of the analysis indicates that the performance of the 120 Hz refresh rate classification is slightly better than the 75 Hz refresh rate. The variance is substantial, however. Even, a high refresh rate will increase the reliability of the high-frequency flickering stimuli by reducing the distance between the two adjacent frequencies.

2.3.2 The Number of Channels and Electrodes Locations

Many studies have recognized that the performance of SSVEP-based BCI speller vitally depended on the electrode positions. For this reason, some of the research tried different electrode positions while testing their methods, and others just attempted to find the optimal electrode position design. Researchers examined three different channel sets while testing their method and observed that for most of the users, the best set was the one that covers the occipital area [32].
A more recent study completely focused on electrode positions and reducing the electrode number [33]. In this study, only four frequencies were used to tag four boxes that contained Latin alphabet characters along with 'delete', 'back' commands—at least three steps required for a successful spelling. The experiment consisted of 3 phases, and after each phase, the number of electrodes are reduced to 16, 6, 4, respectively. A minimum energy combination is utilized in the processing part. For each subject, the best channels are selected for the next phase. The accuracies and ITRs are (94.61% and 27.50 bit/min), (91.27% and 24.09 bit/min), and (93.22% and 23.23 bit/min) for 16, 6, 4 electrodes respectively.
For some speller systems, electrolyte gel is applied on the skin (mostly on the hairy parts) to receive decent signals. For this reason, the real-life applications of spellers seem harder. A study has focused on this problem and tried to discover if it is possible to receive a feasible signal from the skin's hairless areas [34]. 256 channels are utilized, and electrodes are located on the face, behind ears, and neck areas, along with the ordinary parts. In an offline experiment, the stimulus consists of only five frequencies; However, in an online experiment, 12 frequencies are used for designing the stimuli, and extended CCA is used for classification. As they expected, the occipital area performed much better than the other areas, but behind the ear areas also performed well at long time windows. This study concluded that it is possible to obtain the EEG data from the back of the ears for patients that have to lie down face up.

2.3.3 Signal Processing Methods for Target Identification in SSVEP-based BCI

In recent studies, Canonical Correlation Analysis (CCA) method has been widely applied in many experiments, and it is considered one of the most powerful methods for distinguishing between possible frequencies in the frequency component of SSVEP [35, 36]. CCA is a statistical multivariable tool for calculating the correlation of multidimensional variables between two sets. Yet, better performance than standard CCA was obtained by different approaches.

In one study, a new frequency recognition method called Multivariate Synchronization Index (MSI) has been developed [37]. This approach is based on the S-estimator, a nonlinear dynamic theory algorithm that estimates the synchronization of EEGs and reference signals. The results show that the MSI has higher accuracy in shorter data length and fewer channels. Out of three frequency recognition methods (CCA, MEC, MSI), MSI is the most reliable than the other two methods. Furthermore, MSI has been extended in the same study as EMSI. In this method, during the calculation of the synchronization index, the time delayed version of the EEG data was incorporated. Results showed that the extended method outperformed the previous one with an ITR rate of 49.76 as an average of 11 subjects [38]. Furthermore, another extended MSI approach called Temporally Local MSI (TMSI) has been refined by [39]. Because the extraction of information by the TMSI method discriminates by taking advantages of the temporally local EEG signals structure, the classification accuracy has improved for different time windows than standard MSI. Another approach compared with CCA is named Multiset Canonical Correlation Analysis (Mset CCA). The purpose of this approach is to refine the reference signal generated from standard features, and this optimized reference signal is focused solely on training data. The Mset CCA improves the efficiency of frequency recognition, particularly for short data lengths and for a small number of channels compared to the CCA [40].

Based on the combination of two previous methods, CCA and Mset CCA, a new method is created and it is called Multilayer Correlation Maximization (MCM). Therefore, it incorporates the strengths of both approaches. Using three layers of [41] correlation, the MCM approach can obtain frequency information and regular features. This research shows that, compared to using fewer layers, using the three correlation layers method contributes to the highest accuracy.

Another study is based on the CCA method. In this study [42], the training data is involved in the CCA reference signal instead of only using an artificial reference signal that consists of a sin-cos wave. In other words, the principle of incorporating features from training data into the reference signal has been suggested. Consequently, the frequency detection for SSVEP-based BCI can be effectively improved by integrating individual SSVEP training data.

Task-related component analysis (TRCA) was one of the spatial filtering methods that improves the steady-state visual-evoked potentials (SSVEPs) for a high-speed brain speller. By removing the background electroencephalographic (EEG) activities, this approach promotes the increase in the signal-to-noise ratio [43]—the experimental part of this research was broken down into tests that were offline and online. For the offline part, 12 participants have been recorded, while the data were obtained from 20 participants for the online part.Participants were asked to look at a flickering stimulus matrix of 5×8 that includes characters, numbers, and symbols. The stimuli was encoded between each target with a phase difference of 0.35π, although only nine channels were used to record EEG signals from the scalps of the subjects. From 8 Hz to 15.8 Hz, with an interval of 0.2 Hz, the frequency range was chosen. The offline part results show that the TRCA methodology has greatly increased the classification accuracy and ITR relative to the extended CCA method. This study records the highest ITR with 325.33 ± 38.17 bits/min in a cue-guided task.It was the highest ITR in EEG-based BCIs recorded.

As a result, it was sufficient to incorporate phase information to design the visual flickering to increase the number of targets. Besides, another study is called Phase Constrained Canonical Correlation Analysis (p-CCA) [44] has introduced a new idea. This method is based on adding a constraint to the reference signal. The estimated SSVEP response phase is added as a constraint. As a conclusion of this study, adding a constraint to the CCA method leads to enhanced performance. Hence, the accuracy increases by approximately 6.8 %.

3. Problem Formulation and a Summary of the Introduced Approach

This thesis aims to enhance the frequency recognition of SSVEP-based BCI to build a reliable speller system that assists disabled people to interact with their environment. On a computer screen, a speller matrix of 5×8 is shown, and this matrix contains several characters and numbers that flicker repeatedly at different frequencies. The goal is to recognize the target character, i.e., frequency, from N_c stimulus frequencies. Furthermore, we use the SSVEP benchmark dataset in this study [45], and this dataset is composed of 40 flickering targets. Therefore, we are dealing with a 40-class classification problem. Moreover, the dataset is recorded from 64 EEG channels, and also it is collected from 35 subjects. $X \in \mathbb{R}^{C \times N}$, X is a multi-channel EEG signal, C is the number of channels, and N is the number of time samples. In this study, only 9 channels out of 64 channels are used. Given that X is a multi-channel EEG, some preprocessing methods such as band-pass filters are performed to minimize noise and artifacts. Afterward, we obtain the necessary information from the signals by employing a feature extraction method. The number of stimuli equals the number of classes, i.e., $Y_i \in \{1, ..., N_c\}$ are the labels for Z_i, where Z_i is the extracted feature set from EEG signal that corresponds to subject i. We use three different feature extraction methods in this work, Canonical Correlation Analysis (CCA), Power Spectrum Density Analysis (PSDA) via Welch's method, and Correlated Components Analysis (CORRCA) to explore its performance with our framework. ECOC paradigm is investigated to solve the multi-class classification problem. In general, the ECOC framework contains two primary steps: encoding and decoding. In the encoding part, an ECOC coding matrix $M^{N_c \times n}$ is utilized, where N_c denotes the number of classes and n denotes the number of binary classifiers. Hence, in the coding matrix, the rows represent the codeword, and the columns represent the base classifiers. For three or more classes, the ECOC algorithm reduces the classification problem to a n series of binary classification subproblems, where n is the length of the codewords. The length of the codeword can vary, and it depends on the coding matrix design. There are two types of coding matrix, binary matrix that contains two elements $M \in \{1, -1\}^{N_c \times n}$ and ternary matrix $M \in \{1, 0, -1\}^{N_c \times n}$ that includes zero elements as well to ignore some classes during the training process of the base classifiers. The output of all classifiers are combined, and a decoding method is used for class prediction. We use loss-weighted decoding for the decoding part. SVM is one of the most used binary classification algorithms. Therefore, we use it to train our classifiers. Furthermore, we examine the performance of different coding matrix designs of the ECOC framework with three feature extraction methods.

4. Feature Extraction Methods

Feature extraction is the process for dimensionality reduction, and it is one of the main processes that need to be applied to obtain critical information from the data before applying the classification method.
However, selecting the best feature extraction method is challenging; therefore, we choose three different feature extraction methods to evaluate the classification algorithm's efficiency. Recently, there are several feature extraction methods for SSVEP-based BCI. Our current study investigated three feature extraction methods with the ECOC framework; the first feature extraction method is called CCA, and it will be explained in Section 4.1 and the second feature extraction method is named PSD, and it is clarified in Section 4.2, and the last feature extraction method is called CORRCA, and it demonstrates in Section 4.3.

4.1 Canonical Correlation Analysis (CCA)

Canonical correlation analysis is a statistical method used to recognize and describe the relationship between two sets of random vectors [46]. This method is similar to reducing the original signals' dimensionality by accounting for two signals' correlation. The mathematical relationship between two sets is established using the covariance matrices of the corresponding vectors [47]. CCA was introduced in 1936 by [48] to determine the relationship between two sets of variables for instructional research and generalized for more than two sets of variables in [49]. Thus, it represents a general method for obtaining the relationship between two sets of multidimensional data. By considering the correlation of one set of linear combinations of variables and another set of linear combinations of variables, CCA obtains the relation. The objective is to evaluate the linear pair of the highest correlation between combinations [47]. Pairs of linear combinations are referred to as canonical variables, whereas canonical correlations are their correlations. In this way, the strength of the association between two sets of random vectors is measured. By maximization, we focus on a high dimensional relationship between two sets of random vectors in a few pairs of canonical variables. Geometrically CCA measures angles from two linear subspaces, and canonical relations represent cosines of principal angles between the corresponding subspaces [50]. In two signal spaces, X and Y, the CCA seeks instructions in such a way that there is a maximum correlation between the projections following these directions. Consider a CCA on 20 two-dimensional X and Y observations. Arrows represent the desired directions in the original signal space Projections of the sample onto one-dimensional subspaces are presented in the graphs below. The original space is high-dimensional, while the basis of the low-dimensional subspace spanned by the canonical factors determined using CCA are Wx and Wy.

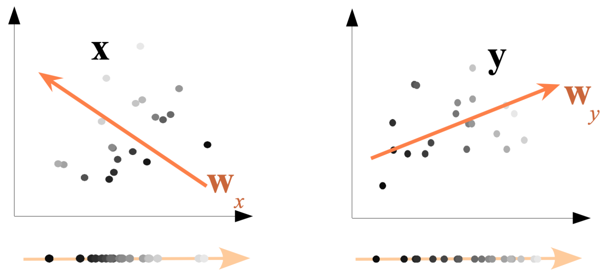

Figure 4.1 Geometric interpretation of CCA [50]

4.1.1 Feature Extraction using CCA for SSVEP-based BCI Speller

In many recent studies, canonical correlation analysis (CCA) has been widely used for frequency recognition in the SSVEP-based BCI framework [35, 36]. Besides, CCA is one of the most efficient methods to discriminate against possible frequencies in the frequency component of SSVEP. Furthermore, the fundamental association between two sets of multidimensional data is calculated by a statistical method. Let's consider X, Y are the two multidimensional variables, where X in the case of SSVEP-based BCI data is the multi-channel EEG data set and Y is corresponding to a set of artificial reference signals and both X and Y have the same length, the following x = $X^T Wx$ and y = $Y^T Wy$ are their linear combinations.
The CCA works to obtain the Wx and Wy weight vectors that optimize the correlation between X and Y by addressing the following formula:

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

(4.1)

$$= \frac{E[W_x^T XY^T W_y]}{\sqrt{E[W_x^T XX^T W_x]E[W_y^T YY^T W_y]}}$$

The maximum of ρ concerning wx and wy is the maximum canonical correlation. The reference signal Y in CCA method is an artificial signal which is generated by sin-cos wave signals due to the fact the SSVEPs signals are characterized at the stimulus frequency and its harmonics by sinusoidal-like waveforms. $Y \in R^{2N_h \times N_s}$ is described as the following:

$$Y_n = \begin{pmatrix} sin(2\pi f_n t) \\ cos(2\pi f_n t) \\ . \\ . \\ . \\ sin(2\pi N_h f_n t) \\ cos(2\pi N_h f_n t) \end{pmatrix}, t = \left[\frac{1}{f_s}, \frac{2}{f_s}, ..., \frac{N_s}{f_s}\right] \tag{4.2}$$

Where, N_h is the number of harmonics, f_s denotes the sampling rate and N_s refers to the time points in each channel.

After finding the correlation between two signals, the highest ρ indicated by the maximum canonical correlation, taking under consideration Wx and Wy, while projections onto Wx and Wy (i.e. x and y) denote the canonical variants, and the output for frequency recognition for standard CCA method is determined by formula 4.3 as clarified in Fig. 4.2 :

$$O = argmax \ \rho_{f_i} \tag{4.3}$$

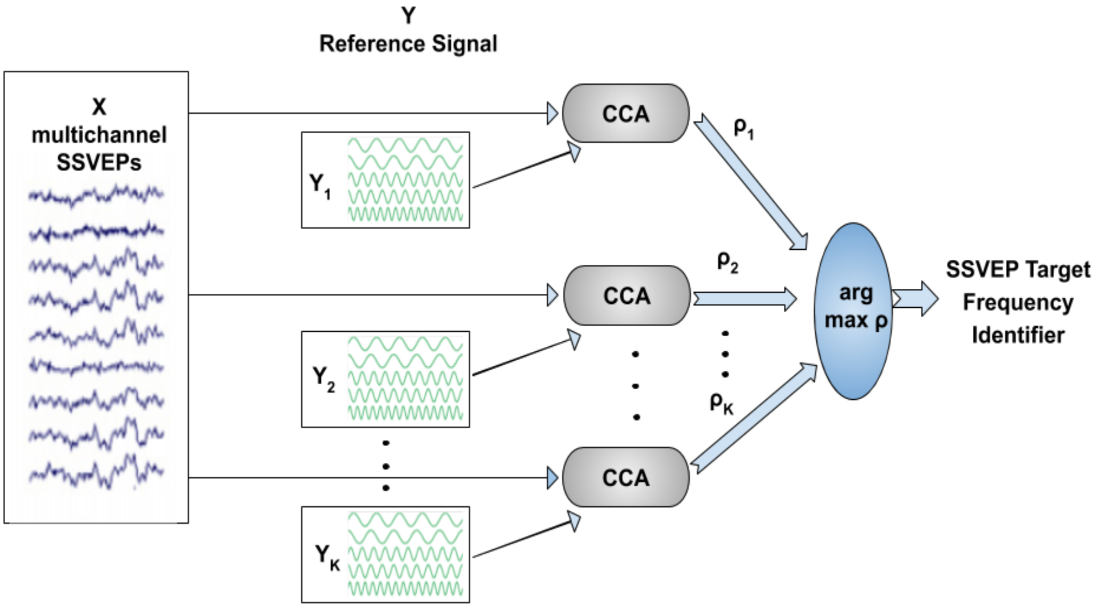

Figure 4.2 An illustration of the standard CCA method for SSVEP frequency recognition. X is a multidimensional SSVEPs signal, Y is the reference signal and K is the number of target stimuli

4.2 Power Spectrum Density Analysis (PSDA)

Periodic signals besides their time dependent intensities can be analyzed by their power spectral density. Spectral analysis aims to decompose the signal into a sum of weighted sinusoids, enabling the analysis of the signal's frequency content. PSD enables observation of the frequency content of $y[n]$ varying with the frequency.

Let us consider discrete time signal $\{y(t); t = 0, \pm 1, \pm 2, \ldots\}$, set of random variables with a mean of zero

$$E\{y(t)\} = 0 \text{ for all } t \tag{4.4}$$

Assuming that $y(t)$ is a second order stationary sequence, its covariance function is defined as

$$r(k) = E\{y(t)\,y^*(t-k)\} \tag{4.5}$$

The power spectral density is defined as

$$\phi(\omega) = \sum_{k=-\infty}^{\infty} r(k)\,e^{-i\omega k} \tag{4.6}$$

PSD is used in many applications to obtain frequency components of a signal for analysis. This method's main advantage is that it allows us to view a signal by its frequency components.

4.2.1 Feature Extraction using PSDA for SSVEP-based BCI Speller

Power spectral density analysis (PSDA) is one of the traditional and popular methods in detecting the desired command in SSVEP-based BCI. It depends on the reality that a periodic sample with an equivalent frequency as the stimulation frequency or its harmonics is derived from the brain signals. Once an SSVEP is present within the brain signals, the frequency domain might be measured with a narrow bandwidth that has been covered from a periodic pattern. Welch's method is a nonparametric method that applies the Fast Fourier Transform (FFT) to estimate the power spectral density (PSD). The welch method consists of three main steps:
- The input data is the EEG signals recorded from brain activity and are divided into N segments (overlapping) that have an equal length.

$$eeg_i[m] = eeg[m+iD], i = 0,...,K-1, m = 0,....,M-1 \tag{4.7}$$

- A window will be applied for each segment, and then the periodogram on each window segment will be calculated.

$$P_i(f) = \frac{1}{NU}\ |\sum_{n=0}^{N-1} w[m].eeg_i[m]e^{-j2\pi fm}\,|^2, i = 0,...,L-1 \tag{4.8}$$

- The estimator of the spectral density will be obtained by averaging the periodograms from N segments.

$$P^W(f) = \frac{1}{N}\sum_{i=0}^{N-1} P_i(f) \tag{4.9}$$

To compute the welch feature, the function "**pwelch**" in MATLAB is used.

4.3 Correlated Components Analysis (CORRCA)

Correlated components analysis (CORRCA) method is similar to the CCA method. CORRCA is produced based on a previous technique called COCA [51], which is based on maximizing the Pearson product-moment correlation coefficient. Hence, CORRCA intends to find the linear components of the data that maximize the correlation coefficient between two multidimensional signals. It creates only one projection vector for the two multidimensional signals, making the difference with the CCA method. CORRCA has been used previously to examine cross-subject synchrony of neural processing [52]. However, in recent studies, it has been used for frequency detection in SSVEP-based BCI [53, 54]. This algorithm's main assumption is that the signal consists of reproducible signal and non-reproducible noise and the directions of the reproducible signal are shared between subjects [55]. CORRCA transforms observed data into components to compute the source of covariation [56]. Let $X \in \mathbb{R}^{C\times N}$ and $Y \in \mathbb{R}^{C\times N}$ be two sets of random vectors where C is the number of channels, and N is the number of time samples. The objective of the algorithm is to find $w \in \mathbb{R}^{C\times 1}$ weight vector such that the linear combinations $x = w^T X$ and $y = w^T Y$ are maximally correlated. That is to obtain maximum correlation coefficient as follows:

$$\begin{aligned}\hat{\rho} &= \arg\max_{w} \frac{x^T y}{\|x\|\,\|y\|} \\ &= \frac{w^T R_{12} w}{\sqrt{w^T R_{11} w}\sqrt{w^T R_{22} w}}\end{aligned} \tag{4.10}$$

where R_{11}, R_{12}, and R_{22} are sample covariance matrices $R_{ij} = \frac{1}{N} X_i X_j^T, i = 1,2$. In order to obtain weight vector w that corresponds to the maximum value $\hat{\rho}$ we differentiate the (4.10) with respect w and set to zero. Assuming that $w^T R_{11} w = w^T R_{22} w$ we obtain the following eigenvalue equation

$$(R_{12} + R_{21})\, w = \lambda \,(R_{11} + R_{22})\, w \tag{4.11}$$

The maximum value of $\hat{\rho}$ corresponds to the principal eigenvector of

$$(R_{11} + R_{22})^{-1}\,(R_{12} + R_{21}) \tag{4.12}$$

Represents the strongest correlation between x and y. The second strongest correlation corresponds to the projection of data matrices corresponding to the second strongest value. Similarly, the highest correlation vector kth is obtained by projecting the data matrices to the strongest eigenvector kth.

4.3.1 Feature Extraction using CORRCA for SSVEP-based BCI Speller

CORRCA is one of the effective strategies for identifying frequencies for BCI-SSVEP. We use the CORRCA method as an alternative method for feature extraction. To adapt the above formula 4.10 to SSVEP-BCI system, we consider the two multi-dimensional signal, X is the training data, and y is the template signals obtained by the average of multiple training trials. Fig. 4.3, shows the standard CORRCA method. Hence, the targets will be chosen based on the maximum correlation as the following:

$$f = max_f \hat{\rho} \tag{4.13}$$

Fig. 4.4, clarifies the CORRCA extracted features that are used in this study.

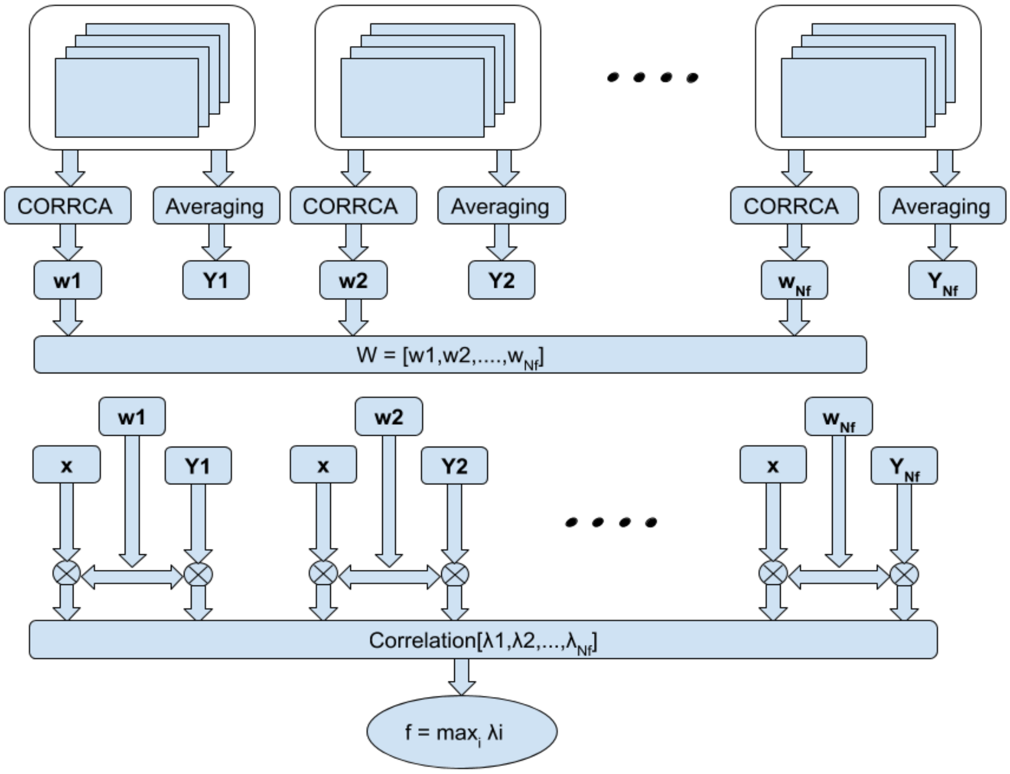

Figure 4.3 Diagram explaining the standard CORRCA method

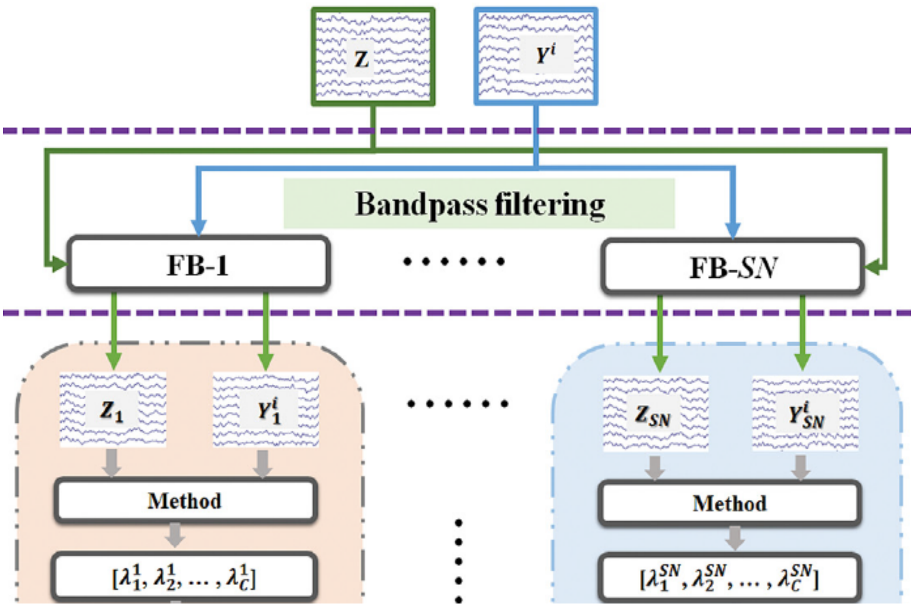

Figure 4.4 Diagram illustrates CORRCA features that are used in this study, where Z is the training data, Y is the template signal, and SN is the number of bandpass filters, taken from [54]

5. Error-Correcting Output Codes (ECOC)

5.1 Introduction to ECOC Framework

Manipulating machine learning algorithms that solve binary problems to distinguish between two classes is usually more manageable than solving a multi-class problem that includes several classes. Some supervised machine learning algorithms are naturally designed to manage a multi-class classification, such as Decision Tree, and Naive Bayes [57, 58]. On the other hand, some algorithms, such as Adaboost and Support Vector Machine (SVM), cannot easily convert into multi-class problems [59, 60]. Besides, many machine learning techniques have focused on solving only binary problems. However, most real-world applications are more complex than having only two classes or labels. In other words, they require to map the input into the corresponding class out of several classes. Researchers and developers over the last few years aim to extend the binary classifier problem to multi-classifiers. One of the effective methods that deals with multi-class classification are Error-Correcting Output Codes (ECOC).

ECOC method has been introduced by Zhang [58], and it is one of the ensemble methods that handles multi-class classification problems. In particular, the essence of this method is combining several binary classifiers to solve a multi-class problem. ECOC framework consists of two fundamental parts: encoding and decoding [61, 62]. The encoding part is based on a coding matrix, where each column in the coding matrix represents a binary classifier, and the rows of this matrix are called codewords; thus, each codeword indicates a class. There are several designs for the coding matrix. Those matrices can differ in the number of classifiers that will be trained, and also the distribution of the elements in each coding matrix can vary. There are two primary types of the coding matrix: binary coding and ternary coding [61]. In binary coding, the coding matrix consists of two elements $M \in \{1, -1\}$, while in ternary coding, three elements are used to design the matrix $M \in \{-1, 0, 1\}$. In this coding matrix design, zero elements are added in order to ignore some classes during the training. Fig. 5.1 shows the ensemble method framework. Several base classifiers will be trained, and the output vector is the combination of the base classifiers outputs. Then, in the decoding phase, we compare the output vector with the coding matrix codewords to find the closest codeword. This framework enables correcting some mistakes of the base classifiers. To clarify, Fig. 5.2, gives a brief example of the error-correcting; the output vector is classified to class c_2, and thus it corrects the mistake of the fourth base classifier D_4. Moreover, there are numerous strategies for finding the closest codeword, like Hamming distance, Euclidean distance, etc. We use loss-weighted decoding in this study, which is also suitable for ternary coding.

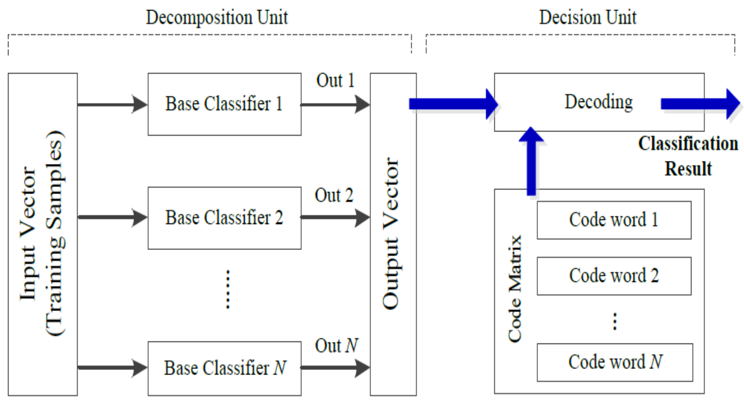

Figure 5.1 ECOC framework for multi-classification tasks, taken from [63]

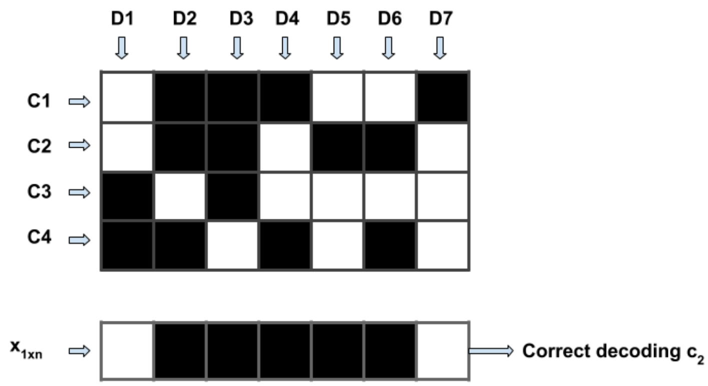

Figure 5.2 Example of the error-correcting, the output vector is classified to class c_2

5.2 Coding Matrix Designs

The coding matrix is used as the encoding stage for the ECOC framework. In the literature, there is no definitive decision about which ECOC design needs to be selected. Therefore, in this work, we investigate several coding matrix designs to evaluate and compare their performance using the SSVEP benchmark dataset. We utilize the most well-known coding matrix designs in this study, one-vs-all (OVA), one-vs-one (OVO), random dense, and random sparse. These strategies are divided into two sections: binary coding and ternary coding. One-vs-all (OVA) and random dense are binary codes that include only two elements (-1,1), and in binary coding, the number of the classifier is usually less than ternary coding, and one-vs-one (OVO) and random sparse are ternary coding that includes "0" elements.

5.2.1 One-vs-all (OVA)

One-vs-all is one of the conventional ECOC coding matrices. In this matrix, the rows are the codewords, and the columns are the classifiers. In each classifier, there is only one class positive and others are negative classes. Let N_m be the number of classifiers in the coding matrix, and N_c is the number of classes for the given classification problem. In one-vs-all (OVA), one class is considered a positive class (+1) while the others are negative (-1), as shown in Fig. 5.3. The coding matrix $M \in \{-1,1\}^{N_c \times N_M}$. The number of classifiers in this matrix design equals the number of classes.

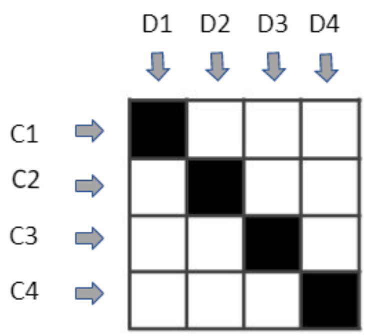

Figure 5.3 One-vs-all ECOC design for a 4-class problem, the black regions coded by 1 and the white regions to -1

5.2.2 One-vs-one (OVO)

One-vs-one coding matrix is a combination of several binary classifiers such that in each classifier, there are three elements (-1, 0, +1), one class is positive while others are negative classes, and some classes will take a value "0" means it will be ignored. Let N_m be the number of classifiers in the coding matrix and N_c represents the number of classes. The coding matrix $M \in \{-1,0,1\}^{N_c^{\times} N_M}$.
The number of classifiers that are used in this method is equal to $\frac{N_c(N_c-1)}{2}$.

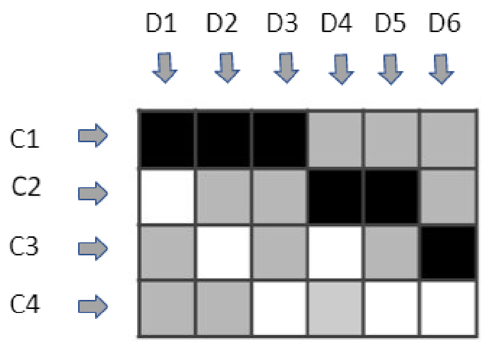

Figure 5.4 One-vs-one ECOC design for a 4-class problem, the black regions coded by 1, the white regions to -1 and the gray position is the 0 symbol

5.2.3 Random Dense

The random dense coding matrix is a matrix that is designed randomly and $M \in \{-1,1\}^{N_c^{\times} N_M}$, N_m is the number of classifiers in the coding matrix and N_c is the number of classes.
The function "*designecoc*" in MATLAB is used to generate the matrix in this study. The software allocates (+1) or (-1) to each element with equal probability of the $N_c \times N_M$ coding matrix, where $N_M \approx 10 \log_2(N_c)$. In this study, there are 40 classes; the number of classifiers in a random dense matrix for 40 classes were 60 classifiers.

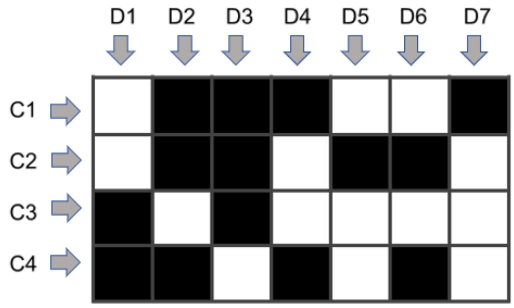

Figure 5.5 Random dense ECOC design for a 4-class problem, the black regions coded by 1 and the white regions to -1

5.2.4 Random Sparse

The random sparse coding matrix is a matrix that is generated randomly and $M \in \{-1,0,1\}^{N_c^{\times} N_M}$, N_m is the number of classifiers in the coding matrix and N_c is the number of classes. The function "*designecoc*" in MATLAB was also used to design the random sparse matrix. The software assigns (+1) and (-1) with equal probability, which is 0.25, and assigns (0) elements with 0.5 probability. The number of classifiers $N_M \approx 15 \log_2(N_c)$. Hence, this matrix design has the highest number of classifiers. In this study, 40 classes are used; the number of classifiers for 40 classes in the random sparse method is 90.

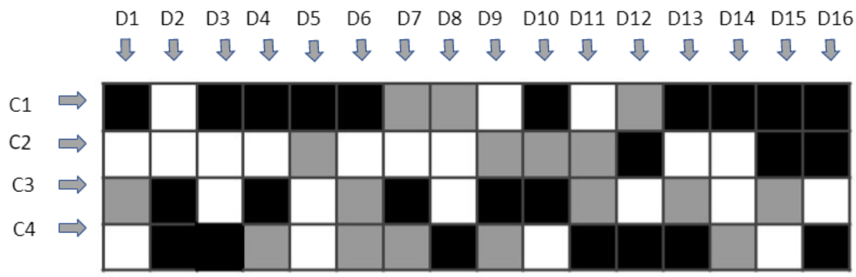

Figure 5.6 Random sparse ECOC design for a 4-class problem, the black regions coded by 1, the white regions to -1 and the gray position is the 0 symbol

5.3.1 Support Vector Machine (SVM)

One of the supervised machine learning algorithms is the support vector machine (SVM) and it is introduced in 1999 [60]. SVM's primary concept is to evaluate the ideal hyperplane that maximizes the margin between two groups. The hyperplane is chosen to separate one class entries from other ones with a maximal margin [64]. Fig. 5.7 shows an example of two-class classification using the SVM method. Moreover, In order to represent patterns in greater dimensions than the dimension of the original feature space, SVM can use a nonlinear mapping platform. Data samples from two distinct classes become separable by a hyperplane for the sake of mapping. [65].
Given a data $\{(\vec{x_i}, y_i), \vec{x_i} \in R^n, y_i \in \{-1, +1\}, i = 1, ..., N\}$. The binary classification can be solved by minimizing the following objective function:

$$min_{w,b,\xi} F = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \xi_i \tag{5.1}$$

$$s.t. y_i(w^T \phi(x_i) + b \geq 1 - \xi_i, i = 1, ..., N, \xi_i \geq 0, i = 1, ..., N \tag{5.2}$$

where ξ is the slack variable and C trades-off margin width and misclassifications.

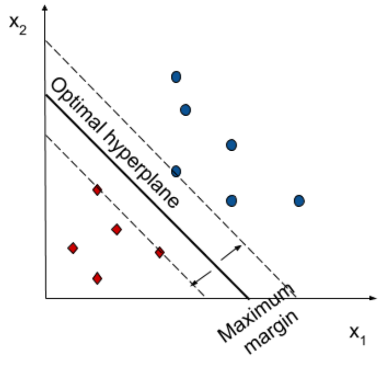

Figure 5.7 Margin and the optimal hyperplane are illustrated for a two-class classification problem on two dimensional (2D) feature space.

In Fig. 5.7, the support vector machines estimate an input-output mapping function from a set of labeled training data. The function is obtained by maximizing the margin between the support vectors of two classes. In addition, several binary learners can be used with the ECOC framework, like KNN and logistic regression. In this study, SVM is used as a binary learner for the ECOC framework.

6. Experimental Results and Discussion

One of the challenging problems is creating an efficient algorithm for the SSVEP-based BCI system to classify the EEG signals to their corresponding stimuli effectively. This thesis demonstrates the ability to design an effective speller system that can provide people with a disability an alternative way of communication. Recently, researchers and developers have focused on enhancing the classification procedure of the SSVEP-based BCI system. Thus, some SSVEP-based BCI applications, like spellers have many flickering targets and it challenging to deal with multi-class classification. In the literature, some supervised techniques were used, like standard CCA and CORRCA. This study uses an ensemble method called Error-Correcting Output Codes (ECOC) to handle the multi-class classification. Furthermore, three different feature techniques were used to evaluate the performance of the classification ensemble method. In addition, for performance measurements, both accuracy and information transfer rate are reported. Four coding matrix designs of ECOC structure are used. This chapter will provide the results of applying the ECOC ensemble methods with different feature extraction methods.

6.1 Dataset

To evaluate the performance of the ECOC ensemble method with SSVEP-based BCI. We choose a publicly available SSVEP benchmark dataset [45]. This dataset can provide us with reliable measurements since it collects it from 35 subjects compared to other datasets that use a lower number of participants. Furthermore, to test the multi-class classification algorithm, the data includes 40 flickering targets.

6.1.1 SSVEP Benchmark Dataset

Over the last few years, the SSVEP benchmark dataset has been widely used in several experiments. Some results showed that this dataset was collected efficiently to meet the requirements for various experimental tests [66, 67, 68]. One of its advantages is that this data includes a high number of stimuli (40 stimuli). Thus, it can provide reliable measurements. Moreover, this dataset is recorded from 35 healthy subjects (17 females and 18 males). Each subject is placed in front of a monitor that displays a 5×8 matrix of flickering targets. These targets are flashing at different frequencies and the frequency range of [8-15.8Hz] with an interval of 0.2 Hz. Fig. 6.1 shows the flickering targets with their corresponding frequencies. The forty targets are twenty-six English alphabets, ten digits, and four symbols. Sixty-four channels are utilized to record the data, and the experiment contains six blocks for each subject. Each block consists of forty trails for each target. As a visual cue, each trial begins with a red square, and it is displayed on the monitor for 0.5 s. Subjects are requested to shift their gaze to the target as soon as possible during the cue duration. Each trial is recorded within 6 s length. Hence, the stimuli are flickering for 5 s, and between each trail, there is a blank screen for 0.5 s. During the experiments, subjects are asked to avoid eye blinks. Furthermore, the data first is downsampled to 250 Hz, and then a notch filter at 50 Hz is utilized to remove and eliminate the noise of the power-line.

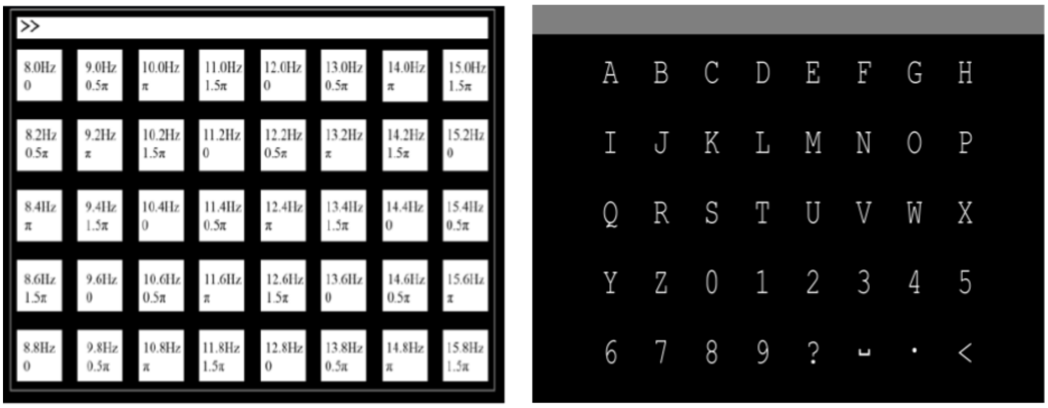

Figure 6.1 Frequency and phase values for all stimuli and their corresponding characters, numbers and symbols

6.1.2 Data Preprocessing

In the SSVEP benchmark dataset, 64 channels were used to collect data and information from the participants. In this study, only 9 channels are selected out of 64 channels, and this selection is based on electrodes location, mainly the electrodes located in the occipital area; this area is responsible for visual processing. These channels' electrode names are Pz, PO5, PO3, POz, PO4, PO6, O1, Oz, and O2. In Fig. 6.2, the electrodes position are highlighted in green.

The data is first downsampled to 250 Hz to reduce the workload and increase data processing speed. Then, with an infinite impulse response (IIR) filter, we apply band-pass filters from 8 Hz to 88 Hz. Using the filtfilt() feature in MATLAB, zero-phase forward and reverse filtering was implemented. In addition, we consider a delay of 140 ms as the subject shifts their gaze towards the stimuli.

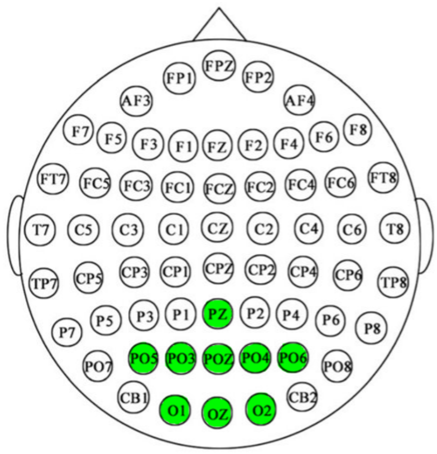

Figure 6.2 The 9 channels that are used in the experiment are highlighted in green color

6.2 Performance Evaluation

The frequency recognition accuracy generally measures BCI-spellers' performance, and to design a reliable speller system, a high-speed system is required.

Accuracy: It shows how the system can be accurate in detecting the desired target frequency and the accuracy is measured by the number of the correct classified target over the total number of target identifications. The percentage of the correct target identifiers, percent accuracy is recorded.

$$Accuracy = \frac{\text{Number of correct target identifications}}{\text{Total number of target identification}} \times 100$$

Information Transfer Rate (ITR): The speed of the system is usually measured by ITR score (i.e. how fast Information may be transferred in one minute in terms of bits). The ITR calculation formula is provided below:

$$ITR \text{ (bits per min)} = [\log(N) + P\log(P) + (1-P)\log\left(\frac{1-P}{N-1}\right)] \times \frac{60}{T}$$

where, N = the number of targets

P = the accuracy of the classification

T = time required to identify one target in seconds

In ITR calculation, we report two different ITR scores. In the first score, we consider the gaze shifting time, while in the other, it is not included in the calculation.

6.3 Analysis of ECOC Designs with Several Feature Extraction Methods

The experimental setup and results are demonstrated in Section 6.3.1 to Section 6.3.3. We explore the ECOC ensemble method with three feature extraction methods CCA, PSDA, and CORRCA.

6.3.1 Analysis of ECOC Structures using CCA Features

Section 4.1 explains the mathematical analysis for the CCA feature extraction method. In this section, we define the essential parameters for the CCA method. The result of applying CCA features with ECOC framework using different coding matrix designs is presented in Section 6.3.1.1. In this work, we employ the standard CCA method to determine the fundamental parameters preliminary. Then, we carry on this result to our framework. One of the main parameters that can affect SSVEP spellers performance is the number of channels. Besides, the SSVEP benchmark dataset is recorded using 64 channels, and some previous studies have selected only 9 channels out of 64, and these channels are located in the central visual area [25, 69].

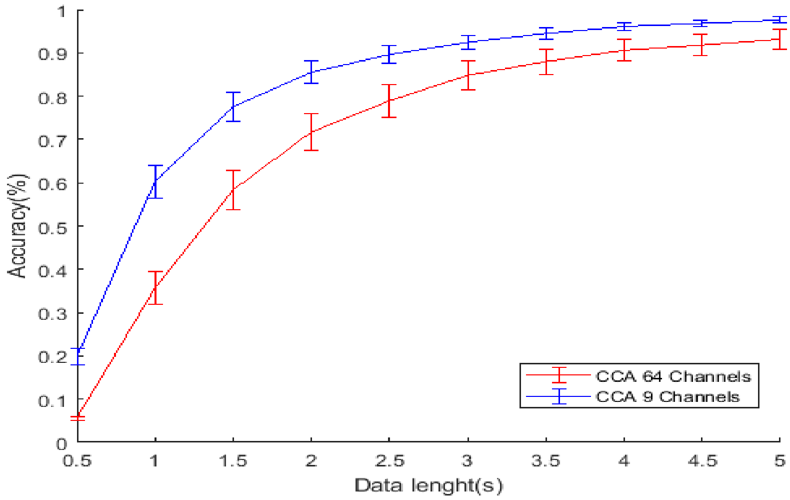

Figure 6.3 A comparison of standard CCA method with 64 channels and 9 channels

Fig. 6.3 shows that using the full set of channels (64 channels) with the standard CCA method reduces performance, especially for a short time window. For example, at 1 s, selecting nine channels improves accuracy by approximately 30%. Furthermore, another critical parameter that can affect the CCA performance is the number of harmonics. The number of harmonics is usually included in the artificial reference signal, as shown in the formula 4.2. In previous studies, the optimal number of harmonics was obscure. Our study reports two numbers of harmonic 2 and 5 because the other number of harmonics like 3 and 4 have a similar performance.
Fig. 6.4 confirms that for a short time data segment from 0.5 s to 1.5 s, using five as the number of harmonic leads to better results. Hence, at 1s, the difference is significant, and it is close to 20% improvement in the accuracy.
In conclusion, based on our experimental results, the CCA parameters are selected as follows: the number of channels is nine, and the number of harmonics for the reference signal is five.

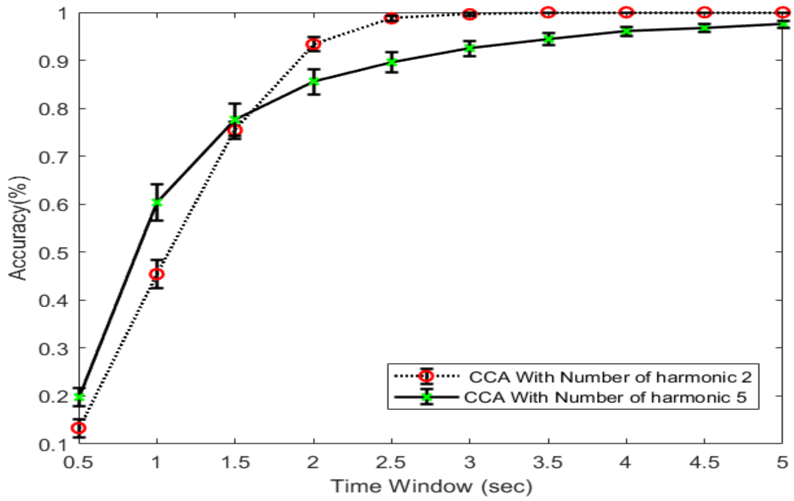

Figure 6.4 A comparison of standard CCA method with two different number of harmonic 2 and 5

Fig. 6.5, illustrates the process of feature extraction using the CCA method for a single subject and one block that is containing 40 trails. The first block in the diagram shows the raw EEG signals using nine channels, and in the second block, a band-pass filter with a range of 8-88 Hz is applied to EEG signals. Then a canonical correlation analysis (CCA) method is performed. The output is a vector of correlated coefficients with a dimension $\{1, 40\}$ for one character/trail, N denotes the number of classes. The final feature matrix for one subject and a single block is a 40×40 matrix.

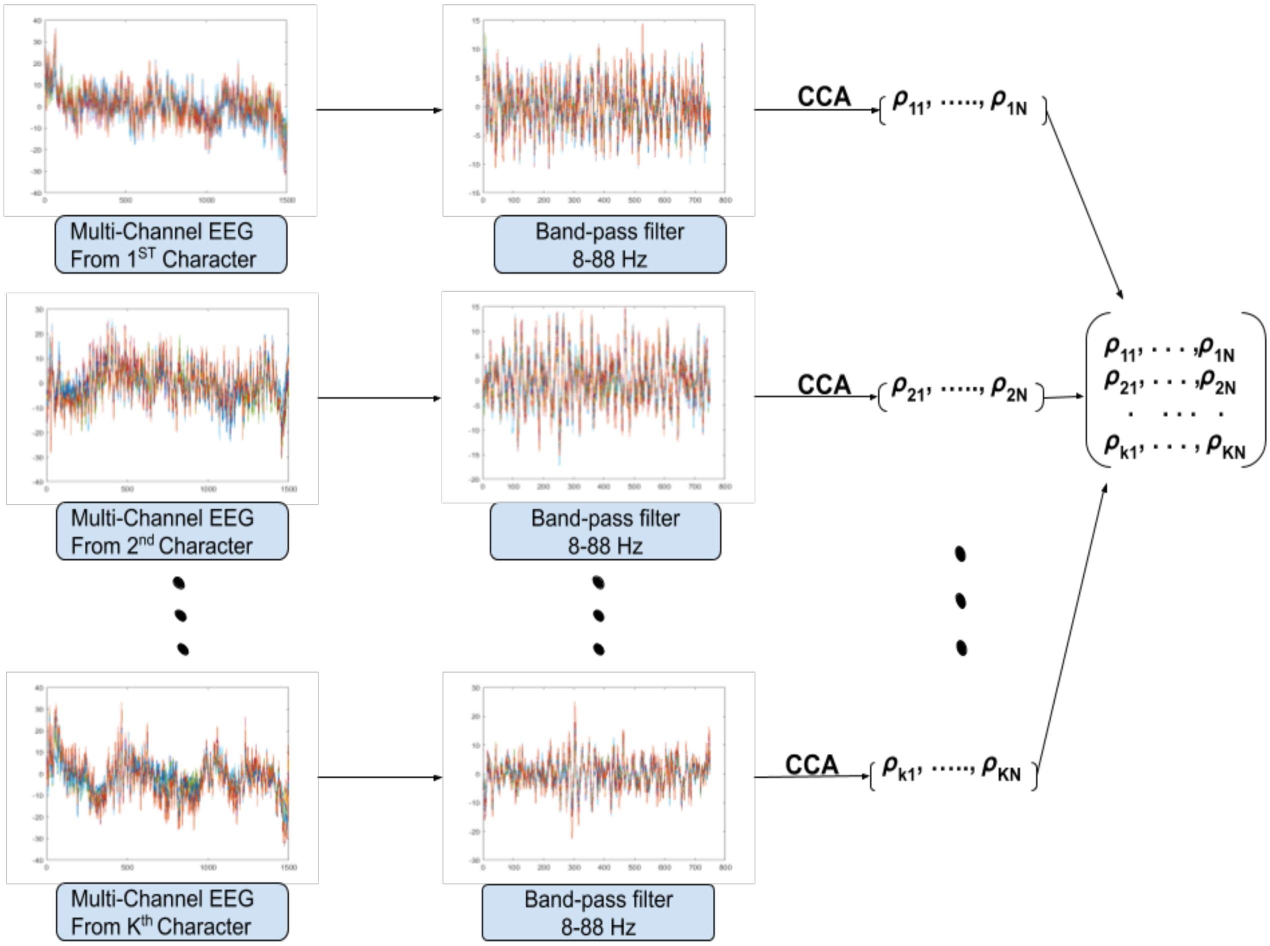

Figure 6.5 A general paradigm for CCA feature extraction steps and the the final feature dimension for one subject one block

6.3.1.1 Evaluation and Results using ECOC Framework with CCA Features

Despite the standard CCA method efficiency, a potential problem is that using the maximum correlation coefficient selection to determine the desired frequency for SSVEP was not adequate to get a high ITR score and classification accuracy especially, for a short time window. To address this problem, this study proposes a novel approach based on an ensemble method to deal with a multi-class classification problem, which is Error-Correcting Output Codes (ECOC). Using CCA features, we apply the ECOC framework for the frequency recognition part. Several coding matrix designs perform the ECOC framework; the four main coding matrices are one-vs-all (OVA), one-vs-one (OVO), random sparse, and random dense. In this study, the four coding matrix structures are used to evaluate the best coding matrix

design performance. In this thesis, the main interest is to get a higher information transformation rate (ITR) for a short time segment compared to a standard feature extraction methods such as standard CCA and standard CORRCA.
In the ECOC framework, each column in the ECOC structure represents a single binary classifier; in our experiments, we use SVM as a base classifier; both Linear SVM and SVM with RBF kernel are integrated into this study. As a result, using a nonlinear SVM with ECOC improves the recognition accuracy compared to linear SVM. Fig. 6.6, shows that using ECOC for frequency recognition can enhance the overall performance compared to the standard CCA method. In both coding matrix design, OVA and OVO using SVM with RBF kernel as a binary learner for ECOC classifiers increase accuracy, especially for a short time window. Meanwhile, for time windows like 4 s and 5 s, the difference is not significant.

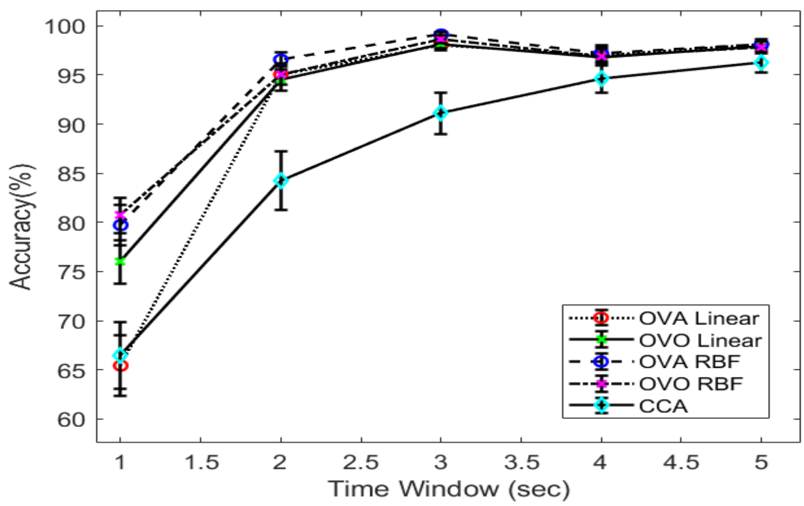

Figure 6.6 The average accuracies across all subjects using Linear SVM and SVM with RBF kernel as base classifier with CCA features from 1 s to 5 s time window with 1s interval

In this part the technical details about CCA feature extraction method with ECOC framework are discussed in details. Furthermore, before applying feature extraction technique, we preprocess the raw EEG signals to reduce the external artifacts in the signals and make it smoother as discussed in Section 4.1.1.
In general, the canonical correlations in the CCA method are bounded between [0 - 1], and the higher value means the correlation between the two variables is high. SSVEP benchmark dataset is gathered from 35 subjects. Each subject can perform differently during the experiments due to several reasons like eye movement, age, etc. Therefore, to evaluate our framework with SSVEP-based BCI speller system, two strategies were used for training the BCI system. The first strategy is described in Section 6.3.1.1.1 and the second in Section 6.3.1.1.2.

6.3.1.1.1 Training Each Subject Individually (Training Per Subject)

In this strategy, the features are extracted from each subject separately since participants can behave differently during the experiments. The dataset that we use consists of 35 participants, eight of them had some previous experience of using the SSVEP-based BCI speller experimental setup from some studies [69, 25]. In contrast, the other 27 subjects were naïve to this kind of experiment. After extracting the features from each subject separately, a single training model is created. We consider that each subject performs individually throughout the experiments, thanks to many reasons like eye movement, age, etc. Therefore, we report the performance of each subject individually and also the average of 35 subjects.
Furthermore, Due to the limited number of features by considering each subject separately, as a data augmentation step, to increase the feature set's size, we use a 50% overlap time window. For example, For 1 s time window, the features are extracted from these time intervals (0-1, 0.5-1.5, 1-2, 1.5-2.5, 2-3, 2.5-3.5, 3-4, 3.5-4.5, 4-5 s). As a result, nine different intervals are used to extract the features for 1 s time window. In the end, these features that are coming from different intervals are combined as one feature set. However, the test set is extracted from the mean interval of the time window. For example, for 1 s, the interval is selected from 2.5-3.5 s time window. Fig. 6.7 describes the general manner of training each subject separately.

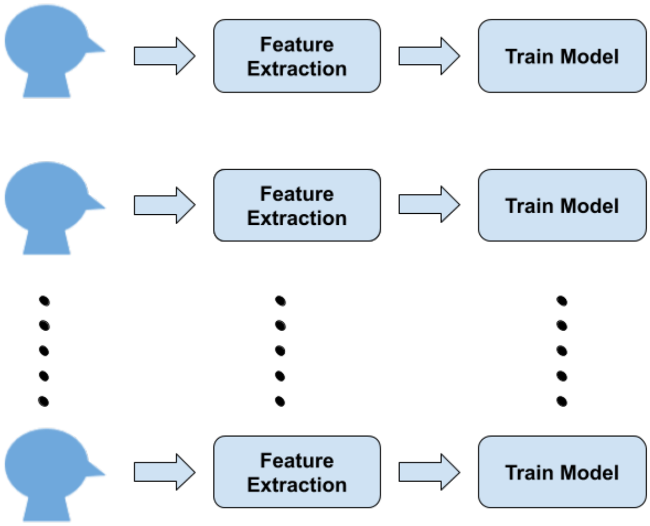

Figure 6.7 The diagram explains the first training method based on extracting the features from each subject separately and creating a single training model for each feature set

6.3.1.1.2 Combining the Features from Overall Subjects with a Single Model Training (Training the Combination of Subjects)

In this training strategy, the features are extracted from each subject, and then they are combined by concatenating it vertically as a single feature set, as shown in Fig. 6.8. For this method, a single model is trained. However, we test this model based on several test sets coming from each subject of an alternative block. The dataset consists of 6 blocks; therefore, we select four blocks for training, one for validation, and the remaining block for testing.

As a conclusion of both training criteria, training per subject method has a better performance than training the combination of overall subjects in the ECOC framework using CCA features.

To clarify, Fig. 6.9 illustrates that training each subject individually and average results across subjects can improve the frequency recognition accuracy for a short time window. Hence, for a short time window like 1 s and 2 s, training each subject individually and then taking the average of them can improve the performance than combining the features from all subjects and using a single model for training. In addition, for the training process, we split the data into three portions, training, testing, and validation. The SSVEP benchmark dataset consisted of 6 blocks for

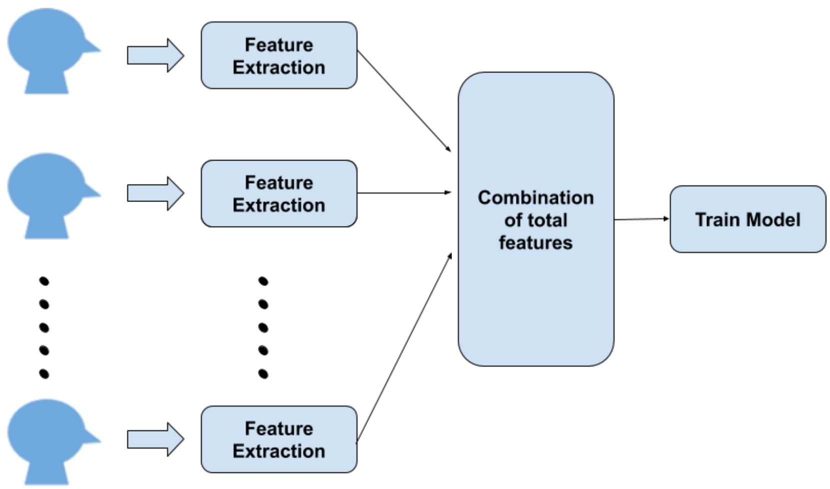

Figure 6.8 The diagram illustrates the second training method that uses the combinations of the features to train a single model

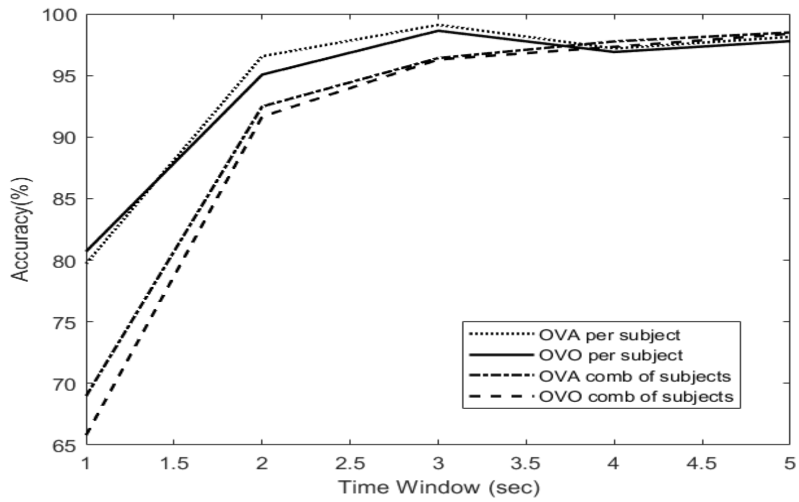

Figure 6.9 A comparison between two ways of training strategies for both ECOC structures OVA and OVO, the first method is training each subject individually and the second method is training the combination of subjects, the CCA features are used with different data lengths from 1 s to 5 s with a step of 1 s

each subject. Each block is containing 40 trials corresponding to all 40 characters. We use four blocks for training, one block as a validation block, and one block for the experiments' test set. Furthermore, A leave-one-out cross-validation is employed to evaluate the classification accuracy. Since four blocks are used for training and one for validation and one block left-out as a test block, to get accurate

results, we repeated this process six times with an alternative test block. In the final result, the average across all blocks are reported.
Moreover, the base classifier for the ECOC framework in this study is the support vector machine (SVM). To determine its optimal parameters, which are C (in case of MATLAB box constraint) and sigma (σ) (MATLAB: kernel scale) in SVM with radial basis function (RBF) kernel, and just C in linear SVM.
Our training strategy in this study is to train a first model and search for the optimal parameters that minimize the loss function by using Bayesian optimization; then after selecting the best parameters C and (σ) in the case of SVM with RBF kernel and C for the linear SVM, the best parameters are utilized to train a new model.
We endeavor to compare several coding matrix designs like OVA, OVO, random sparse, and random dense in this work. As a result, many designs behave similarly, but there is a small significant difference. The number of classifiers is different in each structure, and some coding matrix designs like OVA and random dense are a binary ECOC, while others are ternary ECOC means some classes are ignored in the training process. Fig. 6.10 shows the ECOC structure results for 5 s for all four coding matrix designs.

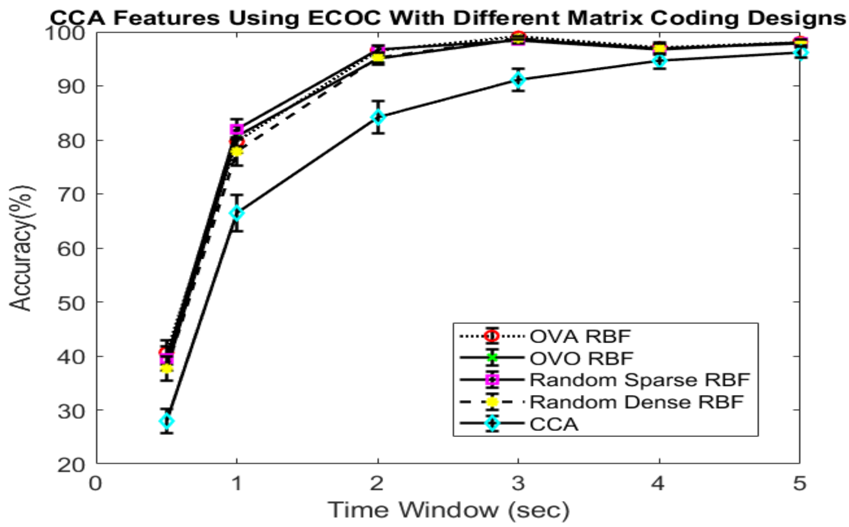

Figure 6.10 (a) Classification accuracies averaged across all subjects obtained by CCA features with ECOC framework for four different ECOC structures OVA, OVO, random dense and random sparse with a SVM RBF kernel as binary learner for a different data lengths from 0.5 s to 5 s

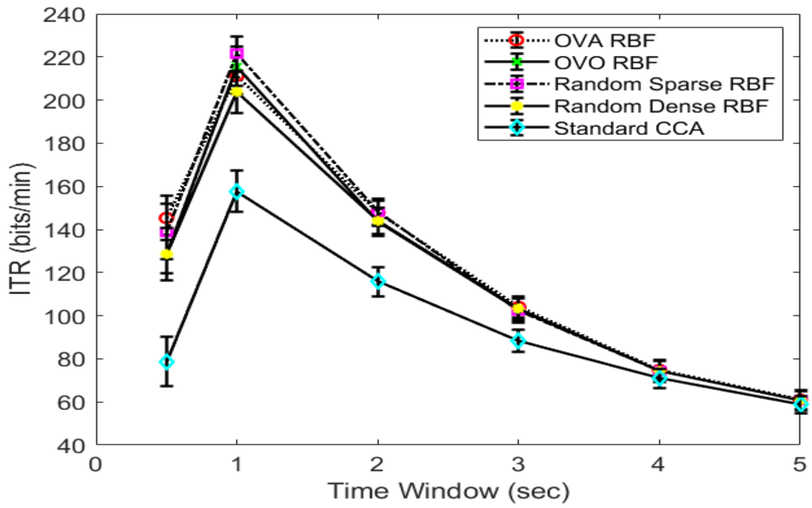

Figure 6.11 ITRs corresponding to the accuracy graph in part(b) when a binary learner is a kernel SVM. The error bars indicate standard errors

Table 6.1 Classification accuracy for 1 s time window per block, each block is the overall average of 35 subjects, and the last row presents the average of the six blocks.

Block Number	OVA	OVO	Random Dense	Random Sparse	CCA
B1	89.30± 11.77	91.45± 10.08	87.68± 11.27	91.58± 10.04	68.35± 19.83
B2	89.27± 15.40	90.67± 11.38	85.35± 20.26	91.90± 9.61	67.14± 23.67
B3	61.45± 19.63	61.05± 19.41	60.70± 20.82	63.04± 19.20	66.21± 22.04
B4	59.32± 17.51	59.98± 16.86	56.11± 19.58	60.82± 17.28	65.28± 18.94
B5	89.85± 11.97	91.14± 10.60	87.98± 12.66	91.84± 9.52	66.14± 21.85
B6	89.11± 10.05	89.92± 9.56	90.11± 11.47	92.97± 9.28	65.35± 21.49
Total Blocks	79.72± 12.32	80.70± 10.87	77.99± 15.47	82.03± 11.15	66.41± 20.04

Table 6.2 Classification accuracy for 3 s time window per block, each block is the overall average of 35 subjects, and the last row presents the average of the six blocks.

Block Number	OVA	OVO	Random Dense	Random Sparse	CCA
B1	99.38± 1.48	98.91± 2.19	98.51± 3.52	98.51± 2.66	92.71± 10.02
B2	99.57± 1.31	98.63± 3.82	99.31± 1.71	99.23± 2.35	90.57± 15.14
B3	99.54± 1.07	99.02± 2.15	99.58± 0.87	99.4± 2.07	91.57± 13.10
B4	99.57± 0.91	98.87± 2.81	98.93± 3.34	98.31± 4.87	91.21± 12.76
B5	99.57± 0.91	99.02± 2.4	98.85± 3.37	98.58± 3.12	89.92± 15.10
B6	97.40± 4.43	97.28± 4.17	97.78± 3.99	96.54± 6.32	90.71± 14.54
Total Blocks	99.10± 1.31	98.62± 2.36	98.83± 1.68	98.43± 3.10	91.12± 12.44

6.3.2 Analysis of ECOC Structures using PSDA features

In this part, another feature extraction method is investigated in our framework. This feature extraction method is based on power spectrum density analysis (PSDA) via Welch method. We use the function "**pwelch**" in MATLAB to compute the welch features. However, some parameters need to be defined, such as the Hamming window and the overlapping. The hamming window in this experiment is selected twice the sampling rate frequency with 50% overlap, and for a short time window like 1 s, the hamming window is equal to the sampling rate.
One of the PSD feature extraction methods' obstacles is that this method is designed for a single channel, but the SSVEP-EEG signal is a multi-channel signal. In this study, we use two approaches to deal with a multi-channel problem for PSD feature extraction. The first approach is explained in Section 6.3.2.0.1 and the second in Section 6.3.2.0.2.

6.3.2.0.1 PSD Computation using Channels' Mean

Based on CCA channel selection results, only nine channels are used with PSD features. However, the PSD algorithm is based on single channels. First approach to overcome the multi-channel EEG signal problem is to apply the algorithm to individual channels and then we take the mean. Fig. 6.12 illustrate the process of computing PSD features by taking the mean of 9 channels.

6.3.2.0.2 PSD Computation using Concatenation of Channels

An alternative way to manipulate the PSD feature with multi-channel EEG signals, is to concatenate the PSD features from 9 channels as one vector. Fig. 6.14 displays the PSD features that are extracted from each channels with their electrode number and the last graph shows how we concatenate them.

6.3.2.1 Evaluation and Results using PSDA via Welch's Method

After computing the PSD features in three different criteria, we calculate the PSD feature for a single channel in the first criteria. However, in the second criteria, we compute the mean of PSDs of 9 channels, and in the last criteria, we calculate the PSDs of 9 channels, and then we concatenate them in the row dimension.

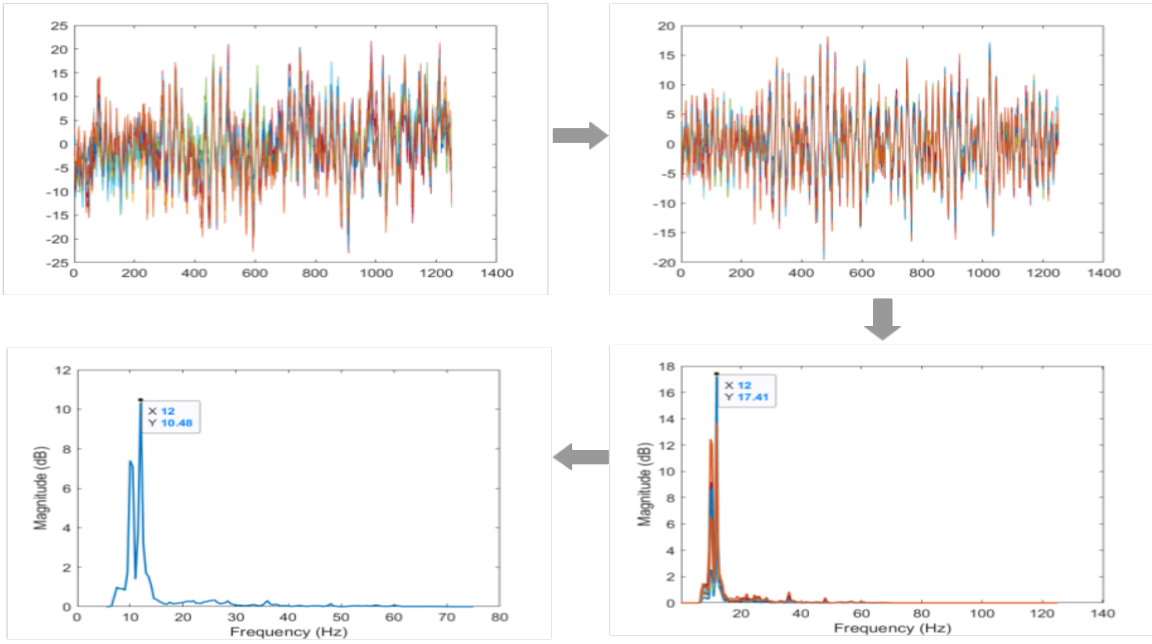

Figure 6.12 The first block shows the raw EEG signals, and then a band-pass filter is applied for the signals, in the next step, the PSD is computed for 9 channels. Finally, the mean of those PSDs is calculated (The diagram presents the fifth flickering stimulus 'E' with 12 Hz frequency)

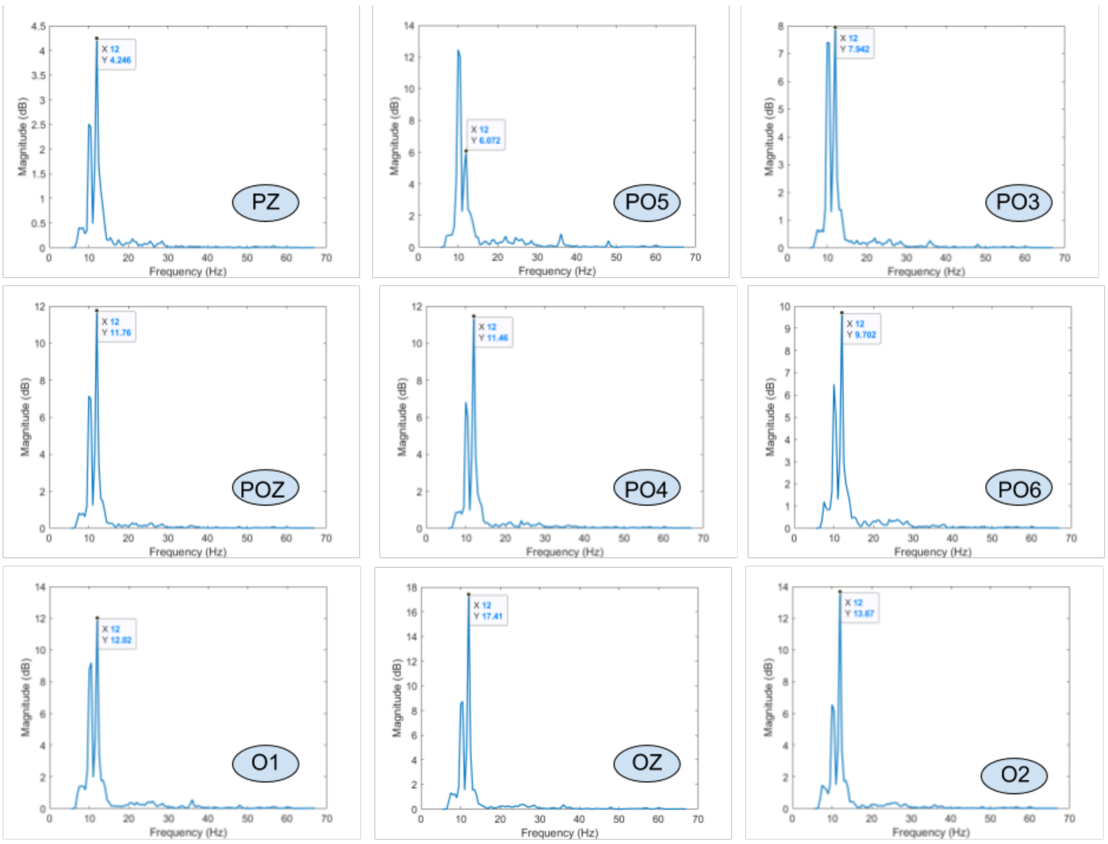

Figure 6.13 PSD feature applied for each nine channels

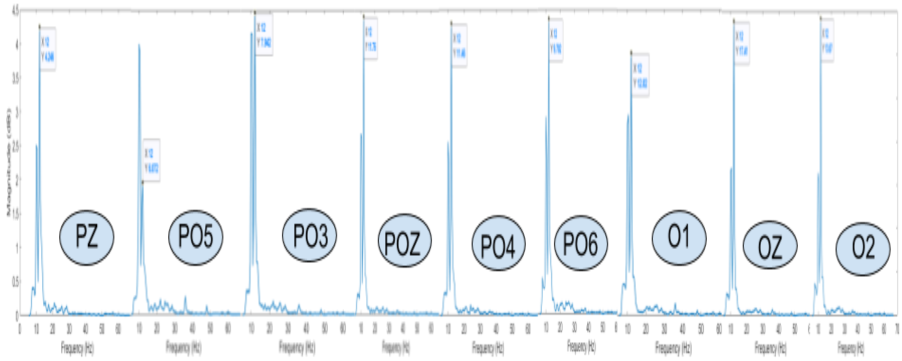

Figure 6.14 Concatenation of PSD features from nine channels

The ECOC framework is utilized to classify these features; in concord with CCA features, the best three coding matrix designs are OVA, OVO, and random sparse. In this work, our goal is to enhance the target identification accuracy as well as the ITR score to establish an SVEP-based BCI speller with high speed and reliable performance. Therefore, both results from a short time window (1 s) and a large time window (5 s) are reported.

The results revealed that the information transfer rate (ITR) when we use PSD features is different from using CCA features. In other words, the ITR score using the CCA feature has only one peak at a short time window (in our experiments, the peak appears at 1 s time window), and then it keeps decreasing along with a higher time window. In contrast, the ITR score using PSD features is fluctuating. Hence, two peaks are developed in our results, the highest peak at 1 s and the next one is located at 3 s time window.

Consequently, the ITR score of CCA feature outperforms the ITR of PSD features.

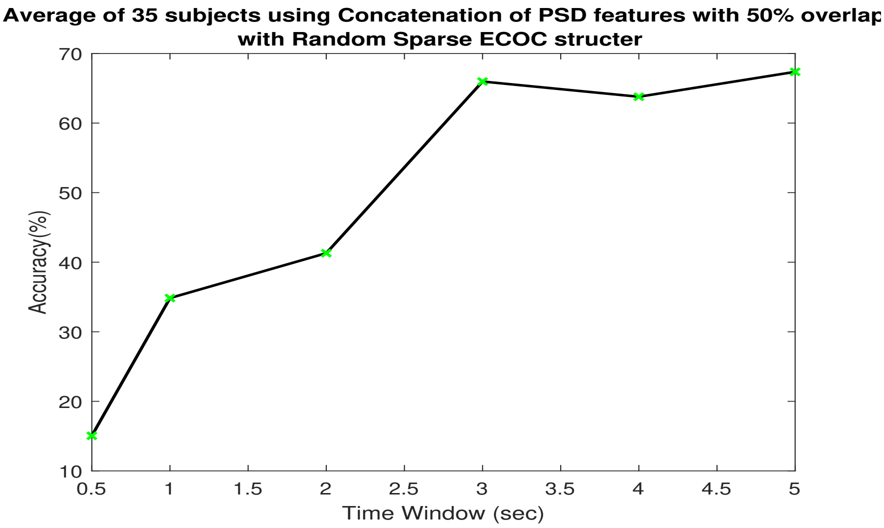

Figure 6.15 Average accuracy across subjects using the concatenation of PSD using random sparse as coding matrix structer

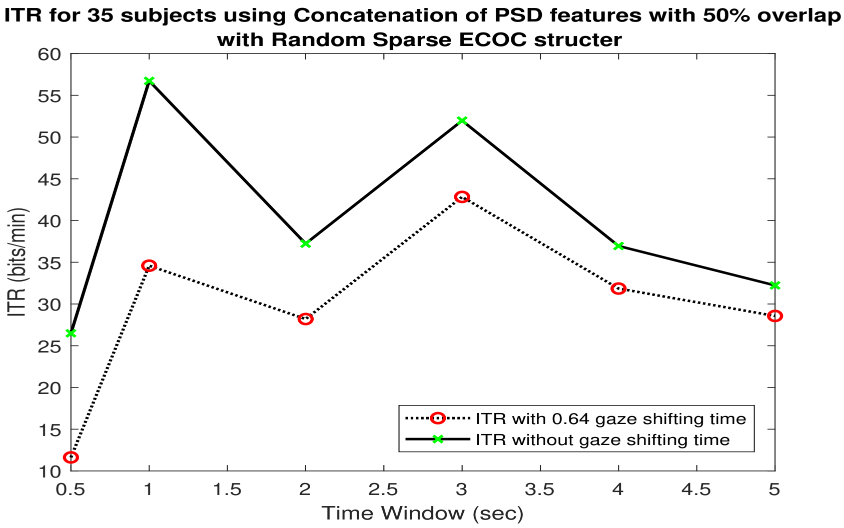

Figure 6.16 ITR score using the concatenation of PSD using random sparse as a coding matrix structure. Hence, two ITR scores are reported with gaze shifting time 0.64 and without considering it

We report the frequency recognition accuracy for three types of PSD features, the first type is calculating the PSD for a single channel. The experimental results show that using the channel that is recorded from electrode "Oz" has the highest performance compared to other channels. However, the concatenation of PSDs with random sparse coding matrix design has the highest accuracy compared to other coding matrix designs. As shown in the table 6.3.

Table 6.3 Frequency recognition accuracy using PSD with OVA and random sparse ECOCs for 5 s.

Electrode Numbers	Electrode Names	Accuracy OVA	Accuracy RSP	Accuracy OVO
48	PZ	49.87	44.38	45.5
54	PO5	57.94	54.98	54.14
55	PO3	68.88	65.67	65.62
56	POz	77.68	75.14	75.01
57	PO4	68.87	64.55	64.79
58	PO6	55.58	50.45	49.52
61	O1	66.46	65.35	63.53
62	Oz	80.04	79.23	78.73
63	O2	67.18	64.55	64.02
Mean of 9 Channels combination of subjects	9 Channels	80.59	80.14	79.44
Concatenation of 9 Channels combination of subjects	9 Channels	83.13	85.20	80.25
Average of 35 subjects using Mean of 9 Channels	9 Channels	64.46	61.27	51.95
Average of 35 subjects using Concatenation	9 Channels	70.78	67.36	65.52

Table 6.4 Frequency recognition accuracy using PSD with OVA and random sparse ECOCs for 1 s.

Electrode Numbers	Electrode Names	Accuracy OVA	Accuracy RSP	Accuracy OVO
48	PZ	8.98	8.84	11.33
54	PO5	10.45	10.59	13.58
55	PO3	16.87	17.33	19.76
56	POz	22.71	22.17	24.54
57	PO4	14.71	13.97	16.90
58	PO6	9.24	9.32	11.66
61	O1	14.45	14.91	16.92
62	Oz	21.17	22.02	23.45
63	O2	13.15	13.32	15.97
Mean of 9 Channels combination of subjects	9 Channels	21.51	22.40	24.9
Concatenation of 9 Channels combination of subjects	9 Channels	25.69	27.20	28.05
Average of 35 subjects using Mean with 50% overlap	9 Channels	21.83	22.20	23.62
Average of 35 subjects using Concatenation with 50% overlap	9 Channels	30.83	34.84	31.178

Table 6.5 ITR score using PSD with OVA and random sparse ECOCs for 1 s.

Electrode Numbers	Electrode Names	ITR OVA	ITR RSP
48	PZ	2.75	2.65
54	PO5	3.86	3.98
55	PO3	10.01	10.51
56	POz	16.97	16.28
57	PO4	7.73	7.00
58	PO6	2.94	3.00
61	O1	7.47	7.12
62	Oz	15.05	16.09
63	O2	6.22	6.38
Mean of 9 Channels combination of subjects	9 Channels	15.44	16.57
Concatenation of 9 Channels combination of subjects	9 Channels	21.31	23.04
Average of 35 subjects using Mean with 50% overlap	9 Channels	15.85	16.32
Average of 35 subjects using Concatenation with 50% overlap	9 Channels	28.34	34.58

Table 6.6 ITR Score using PSD with OVA and random sparse ECOCs for 5 s.

Electrode Numbers	Electrode Names	ITR OVA	ITR RSP
48	PZ	17.79	14.80
54	PO5	25.40	20.74
55	PO3	29.60	27.44
56	POz	35.91	34.03
57	PO4	29.59	26.70
58	PO6	21.09	18.11
61	O1	27.96	27.23
62	Oz	37.72	37.09
63	O2	28.44	26.70
Mean of 9 Channels combination of subjects	9 Channels	38.14	37.79
Concatenation of 9 Channels combination of subjects	9 Channels	39.97	41.86
Average of 35 subjects using Mean	9 Channels	26.64	24.59
Average of 35 subjects using Concatenation	9 Channels	30.91	28.56

6.3.3 Analysis of ECOC Structures using CORRCA Features

CORRCA is one of the efficient methods for frequency recognition for BCI-SSVEP. It has been introduced with BCI paradigm in the following study [70]. In another study [53], the results showed that the standard CORRCA method outperforms the standard CCA method. Therefore, we use the CORRCA method as an alternative method for feature extraction. We investigate this feature with ECOC framework and compare it with other features' performance.

6.3.3.1 Evaluation and Results using ECOC Framework with CORRCA

This section presents the results of applying the ECOC framework with CORRCA features. As an initial step, we compare the performance of standard CORRCA method with CORRCA features with ECOC framework.
Fig. 6.17 displays the average classification accuracy across 35 subjects. The results indicate that using the ensemble ECOC framework can improve accuracy. For example, at 1 s, the accuracy using ECOC structure increase by approximately 10% compared to the standard CORRCA method.

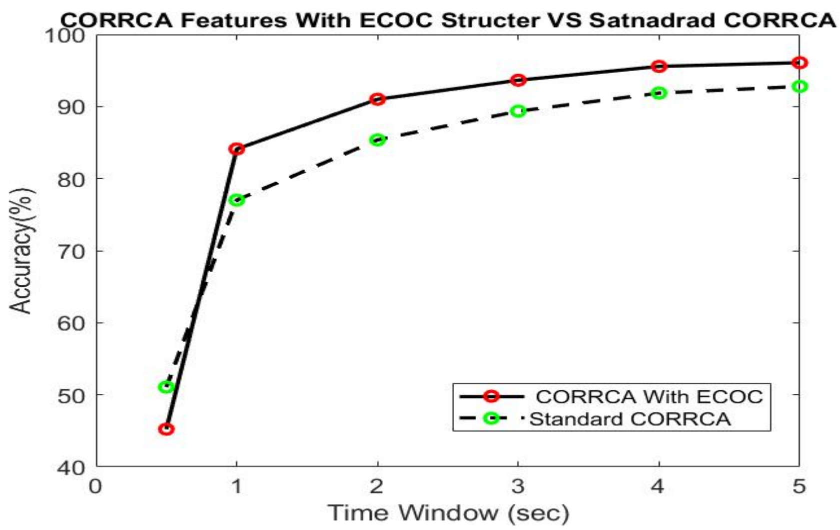

Figure 6.17 A comparison between standard CORRCA and CORRCA feature with ECOC structure for frequency recognition

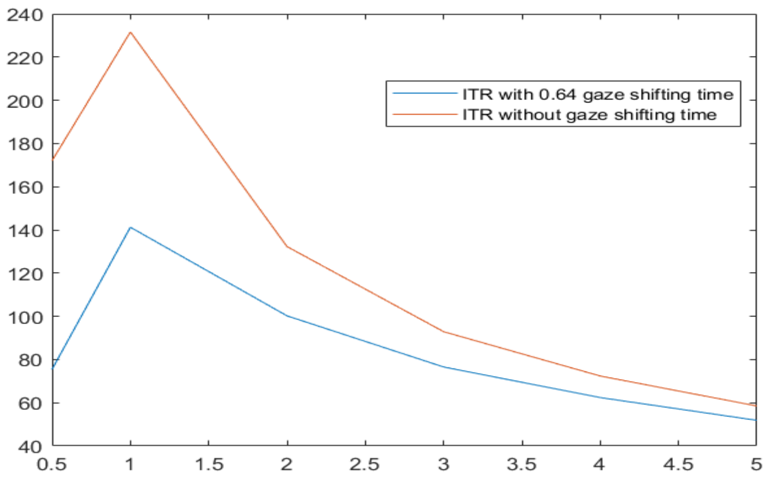

Figure 6.18 The ITR score using CORRCA feature with OVA coding matrix. Two ITR scores are reported with gaze shifting time 0.64 and without considering it

Table 6.7 Accuracy and ITR using CORRCA features with two ECOC structure OVA and random sparse ECOCs.

CORRCA Features	Accuracy OVA	AccuracyRSP	ITR OVA	ITR RSP
1s	84.24	72.61	231.62	181.63
2s	91.00	87.48	132.29	123.48
3s	93.64	90.91	93.436	88.04
4s	95.56	93.367	72.37	69.28
5s	96.075	94.975	58.50	57.22

Table 6.8 Classification accuracy using CORRCA with OVA ECOC. Six time windows (i.e., 0.5, 1, 2, 3, 4 and 5 s) were used for corresponding to the six series.

Subject Number	OVA 05s	OVA 1S	OVA 2S	OVA 3S	OVA 4S	OVA 5S
S1	35	65.875	85.125	76.625	82.875	91.00
S2	36.5	69.75	90.625	92.5	95	96.25
S3	51	79.25	95.125	93.875	95.25	95.13
S4	70.5	92.875	96.625	95.625	96.25	96.63
S5	45	84.125	97.625	98.75	99.75	99.75
S6	50	84.125	94.875	96.75	97.625	98.38
S7	23	64.75	87	93.625	98.875	99.25
S8	30	53.375	64.625	72.75	78.625	83.00
S9	18.5	30.625	46.25	52.5	59.375	69.38
S10	23.5	51.125	64.25	73.375	74.75	77.00
S11	28	64.875	82.5	88.5	89	92.38
S12	68	93.875	99.5	98.375	99.5	99.50
S13	57.5	90.75	96	96.625	96.875	96.88
S14	36.5	76.375	92.125	93	94.75	96.38
S15	31	69.25	87.25	93.875	95.25	95.63
S16	7.5	7.625	14.125	20.375	20.125	29.63
S17	12	22.75	34.125	60.375	60.125	74.50
S18	22	50.625	66	82.125	85.625	92.38
S19	8.5	10.5	16.375	25	29.625	39.38
S20	21.5	39.25	78.25	92.125	94.375	98.00
S21	6.5	15.5	23.25	21.5	19.875	23.88
S22	59	81.125	95.125	97.875	98.5	99.38
S23	23.5	48.375	72.75	71.5	81.75	81.50
S24	16	41.5	71.875	83.875	88.5	94.63
S25	69.5	92.875	99.75	97.75	99.125	99.00
S26	76.5	91.75	99.125	98.625	99	99.00
S27	24.5	52.25	72.625	81.25	83	89.88
S28	24	49.75	74.25	71.75	79	86.13
S29	26.5	50.125	64.75	76	80.5	85.25
S30	27	42.5	66.5	80.375	83.625	90.25
S31	88	97.5	100	99.75	100	100.00
S32	59	86.5	92.75	98	97.375	98.75
S33	3	7.375	12.25	11.875	10.625	9.00
S34	60	90.125	98.625	97.75	98.875	98.50
S35	32	58.875	81.625	87.5	87.75	92.00
Average of 35 subjects	36.3	60.22	74.67	79.2	81.46	84.785
Average of 35 subjects without Bad subjects	40.16	66.67	82.18	86.88	89.38	92.43
Combination of 35 subjects	45.23	84.24	91	93.64	95.56	96.075
Standard CORRCA	51.07	77.012	85.369	89.357	91.869	92.786

7. Conclusions

Recently, researchers have focused on improving the classification procedure of the SSVEP-based BCI system. This thesis investigates an ECOC ensemble method to improve the target identification for SSVEP-based BCI speller. This method deals with multi-class classification problems, and our speller system consists of 40 targets. The results show that the ECOC ensemble method with BCI framework has been executed efficiently to classify the character/target to its corresponding class. We examine these classification methods with some common and conventional feature extraction methods. Those feature extraction methods are PSD via Welch and CCA, and CORRCA. Furthermore, several coding matrix designs for ECOC structures are utilized to evaluate the performance of the ECOC framework. In this thesis, we are interested in improving the information transfer rate (ITR) for a short time window to design a fast and efficient BCI speller. The result showed that using the PSD feature with the ECOC method has the lowest performance than the other feature extraction methods. Hence, PSD's highest ITR score with random sparse ECOC coding design at 1 s is 56 bit/min. Besides, using CCA features outperform the PSD feature with the ECOC ensemble method, and the highest ITR score for CCA feature 222 bits/min for a 1 s time window. Eventually, the CORRCA features with ECOC framework have a better performance than CCA features. The results show that the highest ITR is 231 bits/min at a 1 s time window with an OVA coding matrix. Furthermore, the ECOC ensemble method outperforms both standard method standard CCA and standard CORRCA. Moreover, the performance of the coding matrix designs in terms of the target identification accuracy seems to be similar. However, OVA and random sparse ECOC designs have a slightly more reliable performance than other designs. To conclude, using the ECOC ensemble method for target identification for SSVEP-based BCI can improve the speller performance and build a reliable speller system with a high speed.

References

[1] Masaki Nakanishi et al. “Generating visual flickers for eliciting robust steady-state visual evoked potentials at flexible frequencies using monitor refresh rate”. In: *PloS one* 9.6 (2014), e99235.

[2] Aya Rezeika et al. “Brain–computer interface spellers: A review”. In: *Brain sciences* 8.4 (2018), p. 57.

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

[4] Eric W Sellers, David B Ryan, and Christopher K Hauser. “Noninvasive brain-computer interface enables communication after brainstem stroke”. In: *Science translational medicine* 6.257 (2014), 257re7–257re7.

[5] Shunichi Amari et al. *The handbook of brain theory and neural networks*. MIT press, 2003.

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

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

[8] Marcin Kołodziej et al. “Comparison of EEG signal preprocessing methods for SSVEP recognition”. In: *2016 39th International Conference on Telecommunications and Signal Processing (TSP)*. IEEE. 2016, pp. 340–345.

[9] Paul S Hammon and Virginia R de Sa. “Preprocessing and meta-classification for brain-computer interfaces”. In: *IEEE Transactions on Biomedical Engineering* 54.3 (2007), pp. 518–525.

[10] Zena M Hira and Duncan F Gillies. “A review of feature selection and feature extraction methods applied on microarray data”. In: *Advances in bioinformatics* 2015 (2015).

[11] HS Anupama, NK Cauvery, and GM Lingaraju. “Brain computer interface and its types-a study”. In: *International Journal of Advances in Engineering & Technology* 3.2 (2012), p. 739.

[12] MG Bleichner et al. “Give me a sign: decoding four complex hand gestures based on high-density ECoG”. In: *Brain Structure and Function* 221.1 (2016), pp. 203–216.

[13] Sri Niwas Singh, Fushuan Wen, and Monika Jain. *Advances in System Optimization and Control: Select Proceedings of ICAEDC 2017*. Vol. 509. Springer, 2018.

[14] David Cohen. "Magnetoencephalography: detection of the brain's electrical activity with a superconducting magnetometer". In: *Science* 175.4022 (1972), pp. 664–666.

[15] Ranganatha Sitaram et al. "FMRI brain-computer interface: a tool for neuroscientific research and treatment". In: *Computational intelligence and neuroscience* 2007 (2007).

[16] Shirley M Coyle, Tomás E Ward, and Charles M Markham. "Brain–computer interface using a simplified functional near-infrared spectroscopy system". In: *Journal of neural engineering* 4.3 (2007), p. 219.

[17] Zachary T Harmany, Roummel F Marcia, and Rebecca M Willett. "SPIRAL out of convexity: Sparsity-regularized algorithms for photon-limited imaging". In: *Computational Imaging VIII*. Vol. 7533. International Society for Optics and Photonics. 2010, 75330R.

[18] Christoph Guger et al. "How many people are able to control a P300-based brain–computer interface (BCI)?" In: *Neuroscience letters* 462.1 (2009), pp. 94–98.

[19] Christoph Guger et al. "How many people are able to operate an EEG-based brain-computer interface (BCI)?" In: *IEEE transactions on neural systems and rehabilitation engineering* 11.2 (2003), pp. 145–147.

[20] Soheyl Noachtar and Jan Rémi. "The role of EEG in epilepsy: a critical review". In: *Epilepsy & Behavior* 15.1 (2009), pp. 22–33.

[21] Leigh R Hochberg et al. "Reach and grasp by people with tetraplegia using a neurally controlled robotic arm". In: *Nature* 485.7398 (2012), pp. 372–375.

[22] Jonathan R Wolpaw et al. "An EEG-based brain-computer interface for cursor control". In: *Electroencephalography and clinical neurophysiology* 78.3 (1991), pp. 252–259.

[23] Avery Bedows. *(PDF) Machine Learning Based Epileptic Seizure Detection ...* 2018. URL: `https://www.researchgate.net/publication/339683168_Machine_Learning_Based_Epileptic_Seizure_Detection_for_Responsive_Neurostimulator_System_Optimization`.

[24] Yijun Wang et al. "Brain-computer interfaces based on visual evoked potentials". In: *IEEE Engineering in medicine and biology magazine* 27.5 (2008), pp. 64–71.

[25] Xiaogang Chen et al. "High-speed spelling with a noninvasive brain–computer interface". In: *Proceedings of the national academy of sciences* 112.44 (2015), E6058–E6067.

[26] Ivan Volosyak, Felix Gembler, and Piotr Stawicki. "Age-related differences in SSVEP-based BCI performance". In: *Neurocomputing* 250 (2017), pp. 57–64.

[27] Rafał Kuś et al. "On the quantification of SSVEP frequency responses in human EEG in realistic BCI conditions". In: *PloS one* 8.10 (2013), e77536.

[28] Xiaogang Chen et al. "A novel stimulation method for multi-class SSVEP-BCI using intermodulation frequencies". In: *Journal of neural engineering* 14.2 (2017), p. 026013.

[29] Yangsong Zhang et al. "Multiple frequencies sequential coding for SSVEP-based brain-computer interface". In: *PloS one* 7.3 (2012).

[30] Chuan Jia et al. "Frequency and phase mixed coding in SSVEP-based brain–computer interface". In: *IEEE Transactions on Biomedical Engineering* 58.1 (2010), pp. 200–206.

[31] Yosuke Kimura et al. “SSVEP-based brain–computer interfaces using FSK-modulated visual stimuli”. In: *IEEE Transactions on Biomedical Engineering* 60.10 (2013), pp. 2831–2838.
[32] Min Hye Chang et al. “Eliciting dual-frequency SSVEP using a hybrid SSVEP-P300 BCI”. In: *Journal of neuroscience methods* 258 (2016), pp. 104–113.
[33] Elena Marx, Mihaly Benda, and Ivan Volosyak. “Optimal Electrode Positions for an SSVEP-based BCI”. In: *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. IEEE. 2019, pp. 2731–2736.
[34] Yu-Te Wang et al. “An online brain-computer interface based on SSVEPs measured from non-hair-bearing areas”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25.1 (2016), pp. 14–21.
[35] Zhonglin Lin et al. “Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs”. In: *IEEE transactions on biomedical engineering* 53.12 (2006), pp. 2610–2614.
[36] Guangyu Bin et al. “An online multi-channel SSVEP-based brain–computer interface using a canonical correlation analysis method”. In: *Journal of neural engineering* 6.4 (2009), p. 046002.
[37] Yangsong Zhang et al. “Multivariate synchronization index for frequency recognition of SSVEP-based brain–computer interface”. In: *Journal of neuroscience methods* 221 (2014), pp. 32–40.
[38] Yangsong Zhang et al. “The extension of multivariate synchronization index method for SSVEP-based BCI”. In: *Neurocomputing* 269 (2017), pp. 226–231.
[39] Yangsong Zhang et al. “Robust frequency recognition for SSVEP-based BCI with temporally local multivariate synchronization index”. In: *Cognitive neurodynamics* 10.6 (2016), pp. 505–511.
[40] YU Zhang et al. “Frequency recognition in SSVEP-based BCI using multiset canonical correlation analysis”. In: *International journal of neural systems* 24.04 (2014), p. 1450013.
[41] Yong Jiao et al. “Multilayer correlation maximization for frequency recognition in SSVEP brain-computer interface”. In: *2016 Sixth International Conference on Information Science and Technology (ICIST)*. IEEE. 2016, pp. 31–35.
[42] Yijun Wang et al. “Enhancing detection of steady-state visual evoked potentials using individual training data”. In: *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. 2014, pp. 3037–3040.
[43] Masaki Nakanishi et al. “Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis”. In: *IEEE Transactions on Biomedical Engineering* 65.1 (2017), pp. 104–112.
[44] Jie Pan et al. “Enhancing the classification accuracy of steady-state visual evoked potential-based brain–computer interfaces using phase constrained canonical correlation analysis”. In: *Journal of neural engineering* 8.3 (2011), p. 036027.
[45] Yijun Wang et al. “A benchmark dataset for SSVEP-based brain–computer interfaces”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25.10 (2016), pp. 1746–1752.
[46] Raymond McCollum. “Canonical Correlation Analysis for Longitudinal Data”. PhD thesis. Old Dominion University, 2010.

[47] Richard Arnold Johnson and Dean W. Wichern. *Applied multivariate statistical analysis*. 5. ed. Upper Saddle River, NJ: Prentice Hall, 2002. XVIII, 767. ISBN: 0130925535. URL: http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+330798693&sourceid=fbw_bibsonomy.

[48] Harold Hotelling. "Canonical correlation analysis (cca)". In: *Journal of Educational Psychology* (1935), p. 10.

[49] J. R. KETTENRING. "Canonical analysis of several sets of variables". In: *Biometrika* 58.3 (Dec. 1971), pp. 433–451. ISSN: 0006-3444. DOI: 10.1093/biomet/58.3.433. eprint: https://academic.oup.com/biomet/article-pdf/58/3/433/635083/58-3-433.pdf. URL: https://doi.org/10.1093/biomet/58.3.433.

[50] Michael Reiter. "Enhanced Multiple Output Regression based on Canonical Correlation Analysis with Applications in Computer Vision". PhD thesis. Graz University of Technology, 2010.

[51] Jacek P Dmochowski et al. "Correlated components of ongoing EEG point to emotionally laden attention–a possible marker of engagement?" In: *Frontiers in human neuroscience* 6 (2012), p. 112.

[52] Samantha S Cohen and Lucas C Parra. "Memorable audiovisual narratives synchronize sensory and supramodal neural responses". In: *ENeuro* 3.6 (2016).

[53] Yangsong Zhang et al. "Two-stage frequency recognition method based on correlated component analysis for SSVEP-based BCI". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 26.7 (2018), pp. 1314–1323.

[54] Yangsong Zhang et al. "Hierarchical feature fusion framework for frequency recognition in SSVEP-based BCIs". In: *Neural Networks* 119 (2019), pp. 1–9.

[55] Lucas C Parra, Stefan Haufe, and Jacek P Dmochowski. "Correlated components analysis-extracting reliable dimensions in multivariate data". In: *arXiv preprint arXiv:1801.08881* (2018).

[56] Stefan Haufe et al. "On the interpretation of weight vectors of linear models in multivariate neuroimaging". In: *NeuroImage* 87 (2014), pp. 96–110. DOI: 10.1016/j.neuroimage.2013.10.067. URL: https://doi.org/10.1016/j.neuroimage.2013.10.067.

[57] Dewan Md Farid et al. "Hybrid decision tree and naıve Bayes classifiers for multi-class classification tasks". In: *Expert systems with applications* 41.4 (2014), pp. 1937–1946.

[58] Min-Ling Zhang, José M Peña, and Victor Robles. "Feature selection for multi-label naive Bayes classification". In: *Information Sciences* 179.19 (2009), pp. 3218–3229.

[59] Y Freund. *An adaptive version of the boost by majority algorithm. COLT: Proceedings of the Workshop on Computational Learning Theory.* 1999.

[60] Vladimir N Vapnik. "An overview of statistical learning theory". In: *IEEE transactions on neural networks* 10.5 (1999), pp. 988–999.

[61] Sergio Escalera, Oriol Pujol, and Petia Radeva. "On the decoding process in ternary error-correcting output codes". In: *IEEE transactions on pattern analysis and machine intelligence* 32.1 (2008), pp. 120–134.

[62] Oriol Pujol, Petia Radeva, and Jordi Vitria. "Discriminant ECOC: A heuristic method for application dependent design of error correcting output codes". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28.6 (2006), pp. 1007–1012.

[63] Yuchen Zhou et al. "An intelligent model validation method based on ECOC SVM". In: *Proceedings of the 10th International Conference on Computer Modeling and Simulation*. 2018, pp. 67–71.

[64] Chih-Wei Hsu and Chih-Jen Lin. "A comparison of methods for multiclass support vector machines". In: *IEEE transactions on Neural Networks* 13.2 (2002), pp. 415–425.

[65] S Sathiya Keerthi and Chih-Jen Lin. "Asymptotic behaviors of support vector machines with Gaussian kernel". In: *Neural computation* 15.7 (2003), pp. 1667–1689.

[66] Bingchuan Liu et al. "BETA: A Large Benchmark Database Toward SSVEP-BCI Application". In: *arXiv preprint arXiv:1911.13045* (2019).

[67] Aung Aung Phyo Wai et al. "Towards a Fast Steady-State Visual Evoked Potentials (SSVEP) Brain-Computer Interface (BCI)". In: *arXiv preprint arXiv:2002.01171* (2020).

[68] Kiran Kumar GR and Ramasubba Reddy. "Designing a Sum of Squared Correlations Framework for Enhancing SSVEP-Based BCIs". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 27.10 (2019), pp. 2044–2050.

[69] Xiaogang Chen et al. "Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain–computer interface". In: *Journal of neural engineering* 12.4 (2015), p. 046008.

[70] Yangsong Zhang et al. "Correlated component analysis for enhancing the performance of SSVEP-based brain-computer interface". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 26.5 (2018), pp. 948–956.