Practical BCI Speller based on User-Specific Feature Representation and Convolutional Neural Network

by

Madina Kudaibergenova

Submitted to the Department of Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Computer Science

at the

NAZARBAYEV UNIVERSITY

April 2022

Author .....
Department of Computer Science
Apr 27, 2022

Certified by.....
Min-Ho Lee
Assistant Professor, Department of Computer Science
Thesis Supervisor

Certified by.....
Berdakh Abibullaev
Assistant Professor, Department of Robotics
Thesis Co-supervisor

Accepted by .....
Vassilios D. Tourassis
Dean, School of Engineering and Digital Sciences

Practical BCI Speller based on User-Specific Feature Representation and Convolutional Neural Network

by

Madina Kudaibergenova

Submitted to the Department of Computer Science
on Apr 27, 2022, in partial fulfillment of the
requirements for the degree of
Master of Science in Computer Science

Abstract

Brain-Computer Interface (BCI) systems have a great impact on improving people's lives. One of the popular implementations are BCI spellers which are realized in big sized monitor layouts. However, in real-world such applications are inconvenient and immobile. Therefore, smartphone layouts should be used for real-world BCI spellers as diminutive low-impacted stimuli are suitable to solve such issue. Nevertheless, if the stimuli intensity is diminished it might lead to P300 amplitude reduction, which causes poor classification accuracy. The aim of this thesis research is to propose a state-of-the-art BCI speller with less than 1 mm sized visual stimuli implemented in a smartphone layout. To increase the task performance, subjects were instructed to produce a certain mental task while gazing at the target character. During mental tasks experiments the new event-related potential (ERP) component, the late positive potential (LPP), was discovered between 700-800 ms after the stimuli. The dataset of 14 subjects was collected.

In this study, BCI system is proposed based on classification of the P300 wave which is one of the difficult tasks in processing electroencephalography (EEG) signals as it is affected by surrounded noise and has low signal-to-noise ratio. Before P300 classification, prerpocessing recorded signals is an essential step where some features are extracted and selected. The majority of the previous studies utilizes hand-crafted features to detect P300 wave; however, this technique is inefficient in representing the signals because an environment and subjects vary in each individual experiment. Inspired by this, convolutional neural network (CNN) architecture is suggested for P300 signal detection to extract important features automatically.

Additionally, to characterize the user-specific spatial-temporal features a data-driven optimization approach is utilized. Average spelling accuracy for LDA in individual sequences for passive tasks in normal-speller - 76.3% and in the proposed dot-speller - 78.1%, whereas for active tasks in normal - 94.1% and dot- 96.8%. Average spelling accuracy for CNN results in individual sequences for active tasks in dot-speller before data augmentation EEGNet - 97%, ShallowConvNet - 59%, DeepConvNet - 94%, while after augmentation EEGNet - 97%, ShallowConvNet - 88%,

DeepConvNet - 97%. These results are comparable to traditional BCI spellers. This study represents practicability of creating feasible and useful BCI spellers in the future.

Thesis Supervisor: Min-Ho Lee
Title: Assistant Professor, Department of Computer Science

Thesis Co-supervisor: Berdakh Abibullaev
Title: Assistant Professor, Department of Robotics

Acknowledgments

I would like to express my deep gratitude to thesis lead supervisor **Min-Ho Lee**, Assistant Professor of Computer Science, for patience, guidance, enthusiastic encouragement and thanks for believing in me. Moreover, I am grateful that he took me under his supervision and accepted me as a research assistant.

I would also like to thank thesis committee, including my co-supervisor **Berdakh Abibullaev**, Assistant Professor of Robotics, for their useful critiques of this research work.

And thanks to my **family and friends**, for their advice, support and assistance in keeping my progress on time and less stressful as possible.

This master thesis has been examined by a Committee of the Nazarbayev University School of Engineering and Digital Sciences and the VinUniversity College of Engineering and Computer Science as follows:

Professor Min-Ho Lee .....
Thesis Supervisor
Assistant Professor of Computer Science

Professor Berdakh Abibullaev .....
Thesis Co-Supervisor
Assistant Professor of Robotics

Professor Anh Tu Nguyen .....
Thesis Committee
Assistant Professor of Computer Science

Professor Kok-Seng Wong .....
Thesis Committee
Associate Professor of Computer Science

Professor Jurn-Gyu Park .....
Thesis Committee
Assistant Professor of Computer Science

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Brain-computer interface (BCI) is a system that enables users with disabilities to interact with a computer by brain signals measured via electroencephalogram (EEG) [1, 2, 3]. An EEG uses electrodes, which are attached to a scalp, to analyze and discover brain activity [3, 4]. Moreover, BCI is a non-muscular communication opportunity for people with neuromuscular difficulties. It provides controlling external mechanisms or smart devices (i.e. computer, speech synthesizer, or artificial, robotic arm) by interpreting user intention to action [5]. The main advantages of BCIs are their cheap and easy application and utilization without any risk for brain, and therefore, EEG usage is popular among brain activity researchers [2]. EEG-based devices consist of: cap, electrodes, EEG recorder, visual feedback [1].

One of the popular BCI application among researchers is BCI speller which detects EEG signals to identify the expected character [6]. Furthermore, one of the first published BCI research was the BCI speller which was a great start for further BCI applications. From the recent times, the BCI spellers have been improved, and the main BCI paradigms include P300, steady-state visual evoked potential (SSVEP), and motor imagery (MI) [7]. These categories use individual EEG signal features, and therefore, require to develop specific Graphical User Interfaces (GUIs) for spellers layout [8]. Usually, the BCI spellers represent a combination of letters, numbers, and symbols, and the output is a sentence written by a user. Moreover, there can be an audio output for speech synthesis systems and voice recognition mechanisms [7, 8].

Mainly, BCI researchers emphasis on the process of developing a back-end (a signal processing techniques) of the system to get better performance. However, the first thing user notices is the BCI speller layout, and, therefore, GUI also needs a proper attention.

The motivation is that normal BCI-spellers are large-sized, and therefore, are static and inconvenient [9]. Moreover, stimuli can be unintuitive (i.e. different colors and shapes), which leads to confusing interface [10]. The idea is to create a small-sized BCI speller which would be suitable for real-world application and be applied to a smartphone. The experimental paradigm in this study is designed to propose a user-friendly application that can reduce unavoidable stress caused by imposing external properties (e.g., noisy visual/auditory stimuli).

Additionally, the proposed solution also responds to the BCI speller performance issue: neighbor letters showed similar scores as a target letter, which reduced the classification score. The solution is to change the traditional row-column order flashing of letters in the speller to random flashing of 6 letters, which increased the P300 response (i.e. amplitude visibility) So if a user does not expect when the target letter flashes, it will lead to more visible results.

The purpose of this thesis research is to propose the state-of-the-art BCI speller with diminutive size of stimuli. It is called "Dot-speller", and the main motivation of it is to diminish eye fatigue (i.e. a condition that occurs when eyes become tired from intense use). It minimizes the impact of the external stimulus (e.g., size and intensity) in the BCI speller while demonstrating a comparable performance to earlier studies. The experiments show that a mental task generates endogenous Event-Related Potential (ERP) response improving system performance. In other words, the synthesis of an oddball paradigm (where subjects gaze at repetitive non-target stimuli diversified with less frequent target stimuli) and mental task. A pitch sound is used for imagery task produced by the user when the target stimuli is shown on the screen. Participants were asked to perform two mental tasks: passive concentration (PC) is when the user simply gaze at the target stimuli and active concentration (AC) is when mental task is produced while focusing on the target.

As a result, in EEG signals, a late positive potential (LPP) was detected, which is interpreted as a high level cognitive neural activity that can sufficiently decode the user's intention. For the evaluated dot-speller, an accuracy of 91.9% with an Information Transfer Rate (ITR) of 22.2 [bits/min] is obtained, showing that the approach tested in this article performs equally well as other state-of-the-art ERP systems [34,16,15].

Over the past decade, deep learning (DL) algorithms have achieved high accuracy in fields such as computer vision (e.g. image processing) and natural language processing (e.g. speech recognition) [14–16]. Regarding BCI tasks the most successfully applied techniques are linear algorithms. From the recent time, some of the DL approaches has been utilized in BCI applications [11, 9, 12, 13]. However, the over-fitting issue and irrelevance of features caused by the high dimensionality of EEG data leads to low classification performance of the BCI system. The previous studies [14, 15, 5] suggest channel selection algorithms to reduce the feature dimension, but the channel selection is a time consuming task during training stage. Therefore, no channel selection but instead raw signals are used in the Convolutional Neural Network (CNN) algorithms for P300 classification. Another issue is that conventional machine learning approaches use hand-crafted features which neither represent the signal properly nor address the non-linearity of the signal appropriately. Although the major part of BCI systems depends on the hand-crafted features. An automated feature extraction approach is suggested as a solution to hand-crafted features limitation issue due to its extraction of subject and class dependent features [16]. In this thesis study, CNN is utilized to extract features based on the process of learning data which is applicable to the problem. The CNN models have two convolutional layers (C1 and C2) and a fully-connected (FC) layer. C1 and C2 layers perform the convolution in the temporal and spatial domain, respectively. The model is based on a temporal-spatio features extraction from the given EEG signals. After the training procedure, necessary features are extracted from the FC layer, and Softmax activation function is applied before categorical Cross-Entropy function. Then, the selected features are used to detect P300 wave. The models are tested on the collected dataset. The results of

the experiments demonstrate comparable classification accuracy in comparison to the previous techniques.

This thesis paper is structured as follows: Chapter 2 gives a review of a literature which explains BCI systems in general and focuses on the applicable concepts. Moreover, it includes a brief overview of relevant papers related to the BCI field, providing application of CNN to BCI spellers. Chapter 3 is a full description of methods used during the research, including experimental set, data acquisition, and model explanation. Additionally it provides the detailed information about CNN models architecture, training and testing the models on the dataset. Chapter 4 demonstrates outcome of the solution applied and discusses the results. Finally, Chapter 5 concludes the research while the personal view about the limitations and future opportunities.

1.1 Problem Statement

The main problem of implemented BCI spellers is the big size of visual stimuli (i.e. characters are repetitively flashed) which are inconvenient for applying within smartphone layout, and causes eye fatigue. This thesis research suggests to implement a small-sized BCI speller on the tablet which will be suitable for real-world applications. However, diminishing the size of stimuli leads to lower P300 amplitude and lower system performance. The aim of this research is to present the idea of diminishing the external properties of visual stimuli in BCI speller systems and maintaining the performance in previous studies. Moreover, the suggested mental task (i.e. sound imagery) is performed to elicit endogenous ERP components, and the hybrid-oddball paradigm should improve the classification: passive concentration (PC) and active concentration (AC). To classify measured signals, both linear and non-linear approaches were used: Linear Discriminant Analysis (LDA) and three CNN models. LDA is chosen due to its high performance, and CNNs are utilized to achieve as high performance as LDA.

Chapter 2

Literature Review

In recent times, researchers have been developing a variety of applications that might be useful for people with disabilities and improve their interaction with the world. One of the practical communication applications is the Brain-Computer Interface (BCI) speller [8, 17]. Commonly, the BCI speller provides an opportunity to users to interact with an environment using computer and brain signals, and it is safe and convenient for people with movements difficulties (e.g., 'locked-in' patients). The speller overall consists of monitor which displays certain characters, including letters, numbers, and other symbols, and a recording device. This device monitors brain signals, and the recorded signals further are analyzed by the BCI system [17, 18]. The BCI system is built that it can output the desired letter the user selected and construct the word or a sentence. As a result, BCI spellers allows disabled people to make interaction through their brain signals which are interpreted to words. Generally, the lack of opportunity to directly communicate for such people can affect the life quality, and with the help of BCI spellers they can be independent regaining social life to a comfort level [18].

There are different approaches to record brain signals, and commonly, researchers use electroencephalogram (EEG) to measure brain activity. The EEG uses electrodes, which are attached to a scalp, to analyze and discover brain activity [3, 4]. The electrodes record electrical charges that are resulted from brain activity [8]. EEG is a method to measure signals, and therefore, is widely used in the majority of BCI

systems due to its safe, quick and cheap installation.

The recorded brain signals during BCI experiment are translated to certain commands showing an output as the chosen character on external device (i.e. computer display). Comparing with traditional keyboards with physically selected target keys, in the BCI systems users choose a desired character by simply looking at it and making an intention, and as a result, based on measurements of brain signals and classification points the desired 'key' is selected by the BCI speller system. There are three main BCI paradigms, and this research focuses on one of the most popular among researchers and used by the majority of BCI spellers: Event-Related Potentials (ERP).

2.1 Previous Work

2.1.1 P300 Spellers

Over the years, different BCI spellers have been developed. BCI spellers' graphical user interface (GUI) is also has been changed to achieve better User Experience (UX) [19, 18]. Even though, the most important aspect of BCI system is achieving a high performance and accurate feedback, the GUI should also be observed due to the fact that it is what the user interacts with. The user-friendly design is important factor along with the speller performance. Furthermore, a poor designed GUI of a BCI speller might directly reduce the performance parameters, such as accuracy and Information Transfer Rate (ITR). Based on that, researchers have invented the general design for BCI speller which is a matrix speller, and it is a basic layout for the majority of P300 BCI systems. They have developed it as simple, clear user-friendly BCI speller which is also quite fast and achieves high classification accuracy.

One of the first authors who presented P300 BCI speller are Farwell and Donchin [20]. The authors developed and tested BCI system for communication via computer using the event-related brain potential (ERP) component called P300. There are 36 character (6x6 row-column matrix) presented on a 35-inch CRT screen, and stimuli

size is around 12-14 cm. This paper gave a start for BCI speller research that suggest different BCI paradigms and demonstrate high performance.

Earlier researchers have found that the evoked potentials quality directly connected with visual or auditory characteristics of given target stimuli [21, 22, 23, 24]. Consequently, according to their findings to enhance a BCI system performance certain characteristics of stimuli, such as shape, color, or intensity should be regularized. For instance, [21] suggest to transfer pictures of familiar faces to stimulus, [25] propose to make stimuli randomly flashed, and [22] argue to use distinct colors of stimuli. These studies demonstrate the ways of enhancing the visual or auditory stimuli by intensifying them or making louder, and as a result, evoked ERP response becomes stronger leading to high system performance.

The above mentioned methods to enhance the P300 response, however, has some limitations while implementing in real world situations [26]. Firstly, during experiments users usually get tired of sitting in front of a monitor for a long time, and mostly it depends on tasks complexity, experiment duration, environmental distractions and uncomfortable pose. Secondly, some types of visual stimuli are complicated and difficult to build. For example, in robotics devices, or mobile systems have a LED-typed or a miniature built-in monitor, which cannot be used to demonstrate some visual stimuli, such as faces, different colors, and big sized characters [1, 25, 27, 28, 29].

Furthermore, P300 BCI spellers are usually large-sized which means they are static (i.e. immobile) and inconvenient, and sometimes stimuli can be unintuitive (e.g. different colors and shapes) which leads to confusing interface, so it is critical defect for users, as well as for researchers from other fields, due to their inconvenient system interface. Such BCI systems are not applicable in real-world, and therefore, a smaller stimuli might be more appropriate. Xu et al. [28] demonstrate diminutive asymmetric visual evoked potentials (aVEPs) to improve BCI speller performance. In the following work [27], authors suggest to place unnoticeable visual stimuli outside of the fovea vision and apply the canonical pattern matching method. These approaches demonstrate adequate classification performance of ERP components.

2.1.2 Smartphone-based BCI Applications

Lin et al. [30] argue that the existing matrix BCI speller has two unsolved challenges, namely gaze-dependent and space-dependent concerns. They present a triple RSVP speller that is gaze-independent and space-independent, with an online average accuracy of 0.790 and an online average ITR of 20.259 bit/min. To enhance the ITR, their speller employs the rapid serial visual presentation (RSVP) paradigm, which presents three different characters each time. Each character is presented three times. The stimulus presentation interface is a 90195 pixel rectangle, and the system spells at a speed of 10 seconds per character. As a result, the triple RSVP speller may be included into smart phones (such as smartphones, smart watches, and others).

In a common BCI application, de Vos et al. [31] compared their mobile 14-channel EEG system to a state-of-the-art wired laboratory EEG system. The wireless mobile EEG amplifier performed as well as the wired laboratory EEG system in terms of training and testing accuracies and ITR. The results were as follows: the mobile EEG had a mean online accuracy of 85.8% and the wired EEG had a mean online accuracy of 85.4%. Their approach makes it easier to move BCI applications from the lab to real-world settings, which is one of the biggest obstacles in current BCI research.

Velasco-Álvarez et al. [32] present a BCI system that allows users to operate four popular smartphone messaging apps: WhatsApp, Telegram, e-mail, and short message service (SMS). The well-known visual P300 row-column paradigm (RCP) is used to control the BCI, allowing the user to pick control commands as well as spelling text. The technology delivers synthesized speech commands to the smartphone, which are processed by a virtual assistant operating on the device. The viability of the suggested system is supported by the online performance results as well as the findings of subjective questionnaires.

Obeidat et al. [33] investigated and tested a mobile BCI in a rolling wheelchair that used the edges paradigm on small displays where visual crowding is common. Their mobile edge paradigm outperforms the mobile row-column paradigm in terms of

performance. In this study, all of the advantages of the edges paradigm over the row-column paradigm were found to be true. However, the edges paradigm's reduction in adjacent mistakes was unusually limited to horizontal adjacent errors. The theory goes that the dimensional limits of a smartphone's visual interface design influenced the neurological processes of crowding.

2.1.3 P300 Detection

Generally, the P300 is a component of ERP recodred through EEG. In more detail, it is a positive amplitude in voltage which can be monitored around 300-400 ms in EEG signals [34]. In other words, if a user is given an event (e.g. flashing letter), a specific curve in the signal appears after 300 ms. In P300 BCI spellers EEG electrodes to record brain signals are usually located in parietal lobe. However, Krusienski et al. [29] demonstrated that putting electrodes in the occipital lobe is efficient enough. Furthermore, according to [34] important metrics of brain signal function, which is a part of decision making, are magnitude, topography, time and presence. The P300 can be detected by looking at 300 ms after some event is appeared to the user. In P300 BCI speller, the detection of the P300 highest amplitudes is the most important process. The system performance is directly depends on whether the speller detects the P300 peak waves accurately, which leads to the level of ITR between the subject and the system [34].

In [20] it is mentioned that when the user focuses on a certain character on the speller, the computer can detect desired letter in real time (i.e. online classification). This successful detection is reached by repeatable flashes of rows and columns alternately in the matrix speller. Specifically, during experiments where the user focuses on target letter, if the row or a column contains the chosen element while they are flashed, the P300 peak can be seen after 300 ms, and this ERP component is monitored by the system [20, 34].

Latest research are studying the BCI speller systems based on concentration required (e.g. explicit or implicit). Explicit can be considered as a necessity to move eyes to show an intention to a certain region in the speller, whereas implicit - requires

making a mental task and sometimes even without visually paying attention to a stimuli [29]. During experiments with fixed eyes position, the focus is switched from physical to mental, and this kind of attention is no more connected with eye movement, but with implicit concentration. According to these attention categories, the BCI spellers can be further divided into gaze-independent and -dependent [29, 35]. Currently, the majority of researchers try to implement BCI spellers based on the gaze-independence attention with minimal eye movements, even though commonly all conventional spellers use gaze-dependence attention.

Previous ERP-based BCI studies defined exogenous potentials as the specific brain components that are evoked until about 150 msec after the eliciting stimulus (e.g., P100), whereas endogenous potentials have a longer latency that occurs between 200 to 750 msec after the stimulus onset (e.g., P300) [36, 37]. It was found that an auditory lexical decision task induces a late N400 component [37]. In previous studies, users were instructed either to passively gaze/listen to the presented stimuli or were instructed to actively perform a certain task (e.g., pressing a button, counting) in response to the visual or auditory stimuli. As a result, it has been shown that active tasks, generally, elicit more robust ERP responses [19, 20, 25]. Although this late component has been validated in many typical ERP studies, most BCI applications have focused mainly on the P300 component as it is easily evoked in the oddball paradigm [7, 38, 39].

However, previous studies have provided evidence that diminishing stimulus size can lead to a decrease in the P300 amplitude and therefore to degraded system performance [11]. Diminutive stimuli makes P300 weaker; however, the introduced mental task might help because it produces Late Positive Potential (LPP). It is an ERP component that is modulated by the emotional intensity of a stimulus. On the figure you can see the time when LPP is evoked. Although P300 decreases, LPP increases due to mental tasks with no negative effect on performance.

2.1.4 Classification

Generally, the BCI speller performance is demonstrated through accuracy and the ITR. The first is calculated by the division of a number of correct predictions by the total number of predictions. The second is a combination of the accuracy and a system speed, and the measurement unit is the number of error-free bits per time unit. The ITR can be calculated differently in different BCI spellers: based on commands or letters. Some authors [40] argue that ITR is objective comparison of speller performances among researchers, so it is important to pay more attention to the accuracy. Others [34] claim ITR is essential enough to consider as a system performance measurement, and mention that optimal number of epochs should be around six. Moreover, the number of classes in the P300 BCI speller directly connected to the number of epochs, so that calculating them leads to higher ITR [34].

Recently, researchers have started to apply Machine Learning techniques to improve the accuracy of BCI spellers. Some of the most popular ones are Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) and a variety of Convolutional Neural Networks (CNN). In this section the last classification technique is reviewed.

It is important to mention that it is difficult to compare the strength of suggested classifiers due to different inputs: varied size in relation to a number of recorded channels, a time window size, and an experimental set differs from papers to papers [16].

The Army Research Laboratory (ARL) [41] provides different EEG Models for the researchers. In their EEG Models project there are source codes of CNN approaches for preprocessing and classification of EEG signals. Mostly the models are written in Keras and Tensorflow Python libraries. The aim of the project is to give an access to researchers to successfully utilize EEG-CNN classifiers and make a comparison easier as possible on own dataset.

Lawhern et al. [42] notice that an expected EEG control signal directly impact on extraction features and classification in any BCI paradigm, which causes limitation in application to other individual signal. Generally, CNN is used in computer vision

problems and for recognizing speech, so that it extracts feature automatically and further classifies given input. Regarding EEG classification, CNNs are mainly applied to individual BCI paradigms separately, and therefore, the generalisation issue appears if a researcher applies existing architecture to other BCI paradigms. In the paper [42] authors propose a CNN model design which adequately classifies given EEG signals from any BCI paradigm using simple architecture. They introduce EEGNet, a compact CNN for EEG-based BCI systems. The authors use depthwise and separable convolution layers to build, and make a comparison for within- and cross-subject validation for BCI paradigms, such as P300 visual-evoked potentials (VEPs), error-related negativity responses (ERN), movement-related cortical potentials (MRCP), and sensory motor rhythms (SMR). The researchers demonstrate that EEGNet successfully solves the generalization problem reaches high performance while being limited to dataset.

According to Schirrmeister et al., [43], their CNNs, which are called ConvNets, are a blend of two theories that are beneficial for various learning tasks on natural data like images and audio signals. These signals are frequently structured hierarchically. For instance, typical images contain simple shaped edges that can form complicated shapes. [43] emphasize that ConvNet learns nonlinear features locally using convolutions and the nonlinearity characteristics as well as performance and quality features as combinations of lower-level features which can be achieved by processing several layers. On top of that, ConvNet model employ pooling layers, which result in a rougher feature representation and can improve the ConvNet's translation invariance.

Chapter 3

Methodology

3.1 Experiment Information

The dataset of healthy subjects is used for training and testing different models. The data contains complete record of P300 evoked potentials from 14 subjects (aged 25-33 years, 4 females). The 5 subjects participated in this study were BCI naive and the remaining 10 were BCI experienced in some tasks. All subject who participated in experiments had normal or corrected eyesight and had no mental or neurological issues.

Individuals were thoroughly informed of the trial's aims before to the start of the experiment, and signed confirmation assent was obtained from all subjects. The experiments are approved and certified by the Korea University Hospital Institutional Institutional Review Board (IRB) [1040548-KUIRB-16-159-A-2] and the Nazarbayev University Institutional Research Ethics Committee (IREC).

3.2 Data Acquisition

The subjects were to be sitting in a comfortable chair at a distance of around 80 cm from a 19-inch LCD display with a refresh rate of 60 Hz and a resolution of 1280 1024 during data collection. The smartphone speller was displayed with the use of screen capturing application. EEG data is recorded via BrainAmp EEG

amplifier (Brain Products, Germany) with 32 channels (Fp1-2, F3-4, Fz, F7-8, FC5-6, FC1-2, T7-8, C3-4, Cz, CP1-2, CP5-6, TP9-10, P3-4, P7-8, Pz, PO9-10, O1-2, and Oz). Electrodes with international 10-20 system standard is used along with forehead grounded reference on the nose (Ag/AgCl electrodes with a maximum impedance of 10 kΩ). A notch filter with a sampling rate of 1 kHz and a sample rate of 60 Hz is used to remove DC artifacts from the data. Finally, to filter out noise, a Butterworth filter (5th order) with parameters of 0.5 Hz and 30 Hz could be employed.

3.3 Speller Experiment

3.3.1 Experimental Paradigms and Task Definitions

The experiment is designed to explore the practical perspective as a pixel-level visual stimuli for the row-column speller application in a smartphone environment was utilized. This new speller system is designed within the traditional oddball paradigm but with greatly diminished stimulation (dot-speller).

A 6×6 spelling layout (12 flashes in a single iteration) is used. The target-to-non-target trial ratio is 2:10. Important to mention that the speller experiments' experimental strategy (e.g., procedure, validation) is based on well-established methods in related works [12, 15, 16, 30, 34].

Active Concentration (AC) and Passive Concentration (PC) are two types of mental states that are introduced in this study. Subjects in the PC condition are directed to focus on the presented stimuli without being given any specific mental assignment (e.g. simply gazing at a target). Users in the AC condition are given the task of gazing at the target character and doing a sound imagery task, which must be memorized (8 kHz frequency for 2 minutes) prior to the experiment. The spelling performance of the various mental activities inside the speller layout is basically validated.

The experiments is divided into training and test phases. The classifier parameters are calculated in the training part, whereas the test phase is conducted online, and real-time feedback for the estimated target class is provided to the subject after each

attempt. Note that the ERP systems are basically binary classifications, i.e., AC vs. non-target (NT) or PC vs. NT.

3.3.2 Normal- and Dot-speller Experiments

A normal-speller and a dot-speller are designed and implemented within a smartphone layout. The interfaces are laid out with 36 target visual stimuli ('A'-'Z,'1'-'9,'_'). Individual stimuli are evenly spaced throughout the screen, with six rows and six columns. The stimuli are gray in color, with a black background. In normal-speller (see left side of Figure 3-1), character size is 0.8 cm, and the individual letters are repetitively flashed, as is convention in the speller systems. In dot-speller, tiny dots (less than 1mm) are visually positioned on the top of individual letters, and these dot-symbols are flashed instead of the letter itself (see right side of Figure 3-1). Individual letters of the spellers are flashed repeatedly. The ratio of target to non-target trials is 2:10. Every flash consisted of 6 randomly ordered white characters, as is typical of both spellers [34].

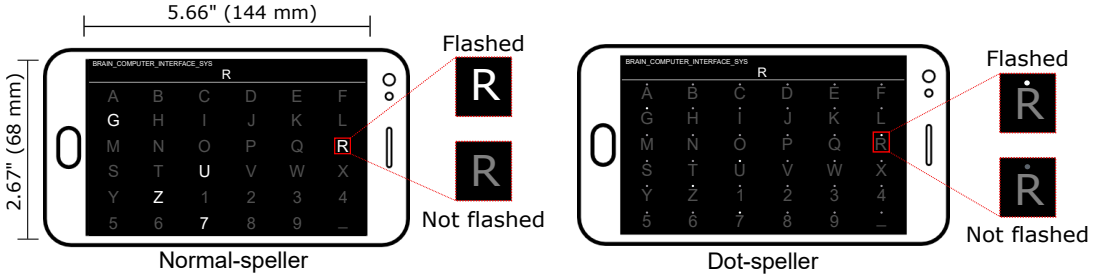

Figure (3-1) **The proposed dot- compared to normal-speller implemented on the smartphone** In the dot-speller (right figure) stimuli is less than 1mm, and are positioned on the top of each symbol in the matrix, and these dot-stimuli are flashed instead of the symbols themselves.

The spellers experiments' experimental strategy (e.g. validation) is based on well-established procedures in relevant studies [3, 11, 12]. A single iteration of flashing letters twice within rows and columns is counted as a sequence of 12 flashes (i.e. trials). There are ten sequences in all, each having a 70 ms stimulus flash and a 150 ms inter-stimulus delay (ISI). Using the screen capture program, the newly created

speller layout was exhibited in a smartphone environment (Galaxy 9, Samsung) with a 1440p OLED/5.8-inch panel.

With the spellers, the experiments are conducted in one session for each subject, which is divided into training and test phases. The user is given two conditional tasks: 1) passive concentration (PC), in which participants are told to glance at the target stimulus while they are idle, and 2) active concentration (AC), in which the target letter is flashed.

For sessions, the experimental technique is essentially the same. The training phase takes place offline, and subjects are given the following sentence to spell: 'BRAIN_COMPUTER_INTERFACE_SYS' (28 characters including spaces '_') according to the task. Hence, 3360 trials (28 characters x 10 sequences x 12 flashes) were collected for individual subject. The training dataset is then used to create two LDA classifiers and two CNN classifiers: PC vs. NT and AC vs. NT.

Subjects are asked to spell throughout the test phase the following sentence: 'U3RQJSMAUWES2QEF KOREAUNIVERSITY' (32 characters with one space '_') according to the given task. In each try, the classifier estimates the target character after showing all the letters (i.e., the end of 10 sequences), and consequently there were 3840 trials (32 characters x 10 sequences x 12 flashes), and this test dataset is used to evaluate the speller system's performance.

3.4 Linear and Non-Linear Classifier Models

Generally, the BCI system starts with data acquisition, then preprocessing, and training CNN model (with automatic feature extraction), testing this model on test dataset, and getting feedback. The BCI speller overall diagram of the proposed method is demonstrated in Figure 3-2. The stimuli are presented on the smartphone P300 speller to the user. The user's brain signals are recorded, and the data of each subject is collected. Then preprocessed signals are given as input to LDA and CNN classifiers. The classifiers give an output feedback with predicted character. Specifically, the dataset is divided into the train and test, and then preprocessed:

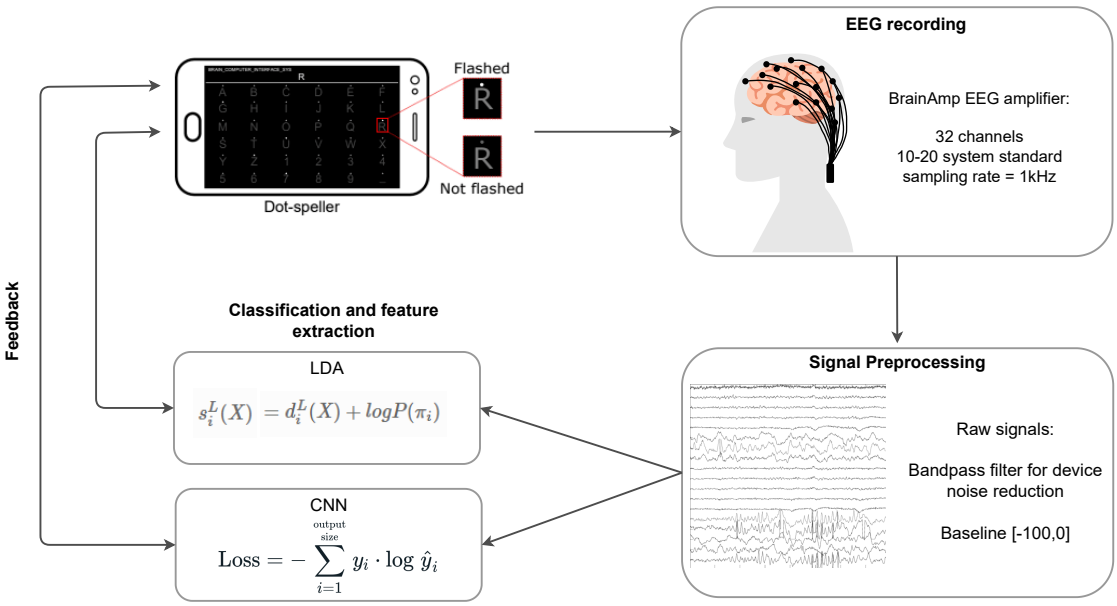

Figure (3-2) **The proposed P300 BCI speller general diagram.** The stimuli are presented on the smartphone P300 speller to the user. The user's brain signals are recorded, and preprocessed signals are given as input to LDA and CNN classifiers. The classifiers give an output feedback with predicted character.

raw signals were extracted from each trial, and users brain signals are detected. The proposed LDA and CNN model are trained using the training set, and the output is fed to test data, which results in prediction classification.

The main reason for choosing CNN models (i.e. EEGNet, DeepConvNet, and ShallowConvNet) is their lightweight character. In other words, for future implementation on smartphone (i.e. without using large personal computers for computations) the lighter model the better. The EEGNet has approximately 10^3 - $2 \cdot 10^3$ training parameters, and the number of them depends on the filters size. The EEGNet architecture consists of 2 Convolutional blocks, and 1 Classification block. The next model, DeepConvNet, consists of 4 blocks, where first block is splitted into two Convolutional blocks to extract temporal-spatial features of EEG input. In comparison to EEGNet it has much more (i.e. $1{,}7 \cdot 10^5$) learning parameters; however, compared to large CNNs, such as AlexNet or GoogleNet, it can be count as lightweight CNN model. The last approach is ShallowConvNet, which is simple version of DeepConvNet because it has same first convolution layers, but has only one block overall. The number of trainable parameters of ShallowConvNet is around 10^5, which is also

counts as lightweight CNN. Therefore, the approaches choice is rational.

3.4.1 EEGNet

Figure 3-3 and Table 3.1, respectively, show a visualization and a detailed description of EEGNet, and the tables in Result section demonstrate the results of the EEGNet model classification for EEG trials recorded at 100 Hz sampling rate with 32 channels and time samples. The Adam optimizer is used to fit the model using standard parameters to minimize the categorical cross-entropy loss function. After a series of tests, the training epochs were shown to be the most effective at 100. Furthermore, the validation is stopped by storing the optimal model weights that result in the smallest validation loss. For simplicity, bias units are not used in any of the convolutional layers, and two-dimensional convolution functions are used instead.

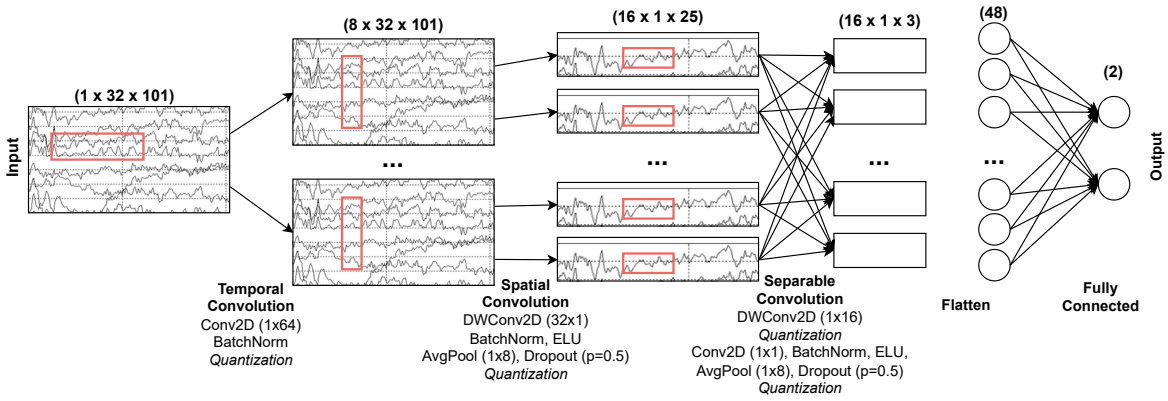

Figure (3-3) **The EEGNet model architecture.** The data is reshaped to (kernel=1, channels=32, sample=101), and the input data is given to Temporal Convolutional Layer with BatchNormalization to extract time-related features over channels. Then they are feed to Spatial Convolution with ELU activation function to extract spatial features over time. Next the Separable Convolution divides a single convolution into several convolutions to produce the same output. Then the features are flattened and passed to Fully Connected layers to classification.

3.4.2 DeepConvNet

The DeepConvNet architecture is built on benchmarks in CNN architectures for computer vision. The EEG signals are changed from top to bottom until they

Table (3.1) **EEGNet model structure.** From the table, the shortened forms mean the following: C=channels, T=time sample, $F1$=temporal filters, D=spatial filters, $F2$=pointwise filters, and N=classes, respectively. Important to mention that, layers options p=0.5 for within-subject and p=0.25 for cross-subject classification are utilized (see section 3.4.1 for more details)

Block	Layer	#filters	Size	#params	Output	Activation	Options
1	Input				(C,T)		
	Reshape				(1,C,T)		
	Conv2D	F1	(1,64)	64*F1	(F1,C,T)	Linear	Mode=same
	BatchNorm			2*F1	(F1,C,T)		
	Depthwise-Conv2D	D*F1	(C,1)	(C*D*F1)	(D*F1,1,T)	Linear	Mode=valid, depth=D, maxnorm=1
	BatchNorm			2*D*F1	(D*F1,1,T)		
	Activation				(D*F1,1,T)	ELU	
	AveragePool2D		(1,4)		(D*F1,1,T//4)		
	Dropout*				(D*F1,1,T//4)		p=0.25 or p=0.5
2	Separable-Conv2D	F2	(1,16)	16*D*F1+ F2*(D*F1)	(F2,1,T//4)	Linear	Mode=same
	BatchNorm			2*F2	(F2,1,T//4)		
	Activation				(F2,1,T//4)	ELU	
	AveragePool2D		(1,8)		(F2,1,T//32)		
	Dropout*				(F2,1,T//32)		p=0.25 or p=0.5
	Flatten				(F2*(T//32))		
Classifier	Dense	N*(F2,T//32)			N	Softmax	Max norm = 0.25

reach the output, as shown in Figure 3-4. Inputs/feature maps are represented by black cuboids, whereas convolution/pooling kernels are represented by brown cuboids. Black and brown, respectively, represent the matching sizes. The sizes listed are for the cropped training version. Weights for all possible pairs of electrodes with filters from the preceding temporal convolution are included in each spatial filter. Note that the proportions of maps and kernels in these schematics are simply estimates.

3.4.3 ShallowConvNet

The model of ShallowConvNet is based on decoding band power features. According to the Figure 3-5 the first two layers perform a temporal convolution and a spatial filter, as in the DeepConvNet (see Figure 3-4). The ShallowConvNet's temporal convolution featured a higher kernel size (25 vs 10) than the DeepConvNet's, allowing a wider range of modifications in this layer (smaller kernel sizes for the ShallowConvNet led to lower accuracies in preliminary experiments on the training set). A squaring

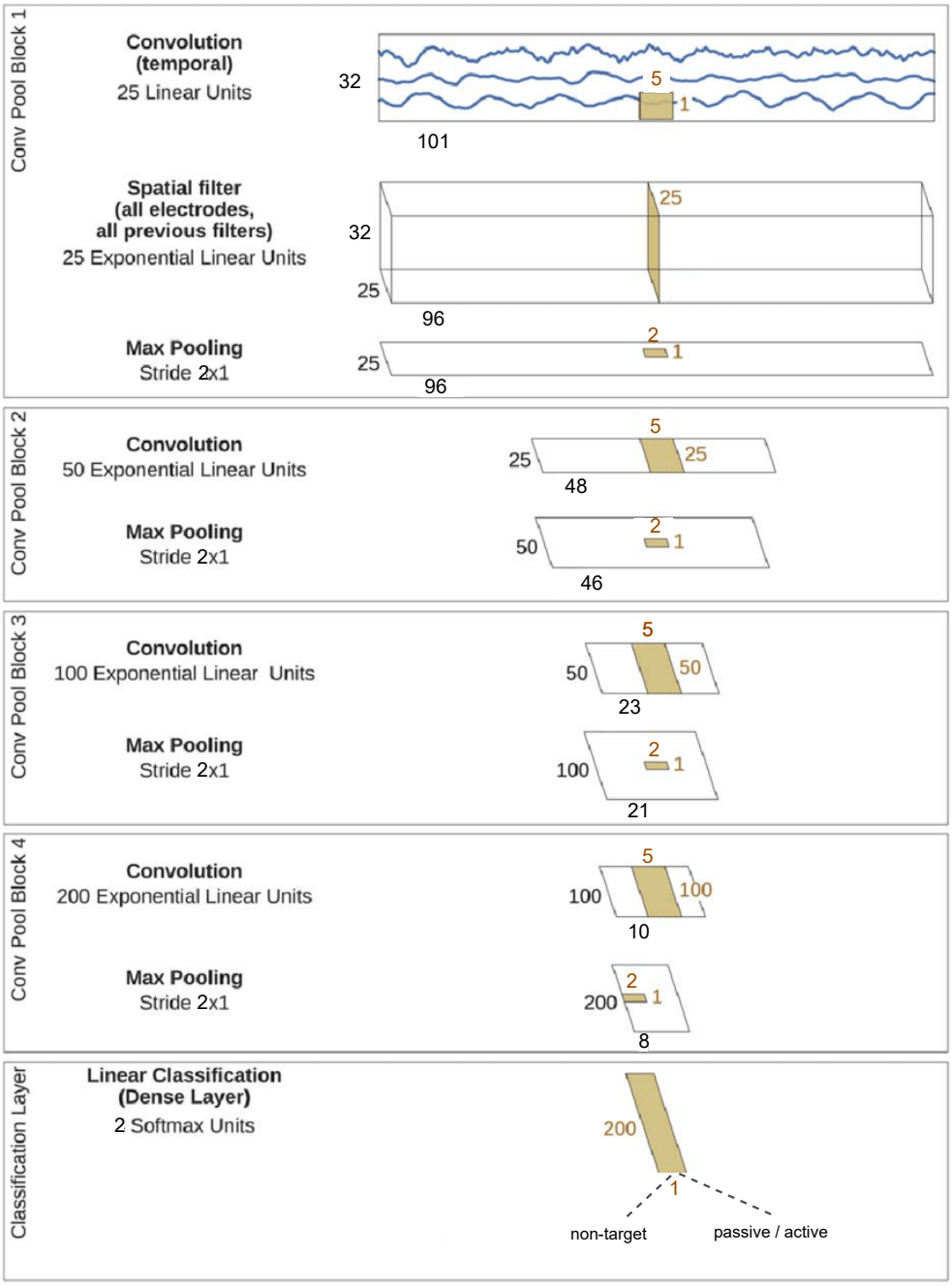

Figure (3-4) **The DeepConvNet model architecture.** The EEG signals are changed from top to bottom until they reach the output. Inputs/feature maps are represented by black cuboids, whereas convolution/pooling kernels are represented by brown cuboids. Black and brown, respectively, represent the matching sizes. The sizes listed are for the cropped training version. Weights for all possible pairs of electrodes with filters from the preceding temporal convolution are included in each spatial filter. Note that the proportions of maps and kernels in these schematics are simply estimates.

nonlinearity, a mean pooling layer, and a logarithmic activation function followed the ShallowConvNet's temporal convolution and spatial filter; however, these stages are not employed in the DeepConvNet. Because the ShallowConvNet encapsulates all computational stages in a single network, all steps can be optimized simultaneously. Also, because each trial has numerous pooling zones, the ShallowConvNet can learn a temporal structure of the band power fluctuations within the trial, as demonstrated.

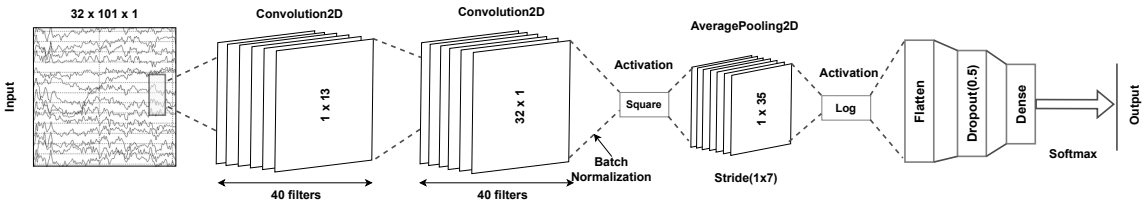

Figure (3-5) **ShallowConvNet model structure.** The data is reshaped as (kernel=1, channels=32,sample=101). First Convolution layer Conv2D extracts temporal features over channels with 40 filters, and the second Conv2D extracts spatial features over time points. After BatchNormalization the Square activation function is applied, and feed to AveragePooling2D layer with stride (1x7), and Log activation function. Features then are flattened with further dropout and dense layer. The softmax function normalized the inputs, and pass them to a classifier.

3.4.4 Tools

The data is recorded and analyzed with hardware and software of Brain Product company. BCI speller layout is implemented on Matlab with the PsychToolbox toolbox. Speller is demonstrated on the smartphone with Multi Monitor Application (i.e. Spacedesk). This application captures the screen of a personal computer and demonstrates the speller on the smartphone. Therefore, EEG signals recording, speller computations, classification were processed on the personal computer, and the smartphone was used as a second screen. The LDA is implemented with Python sklearn library. The CNN model is implemented with Python: brain signals are analyzed and illustrated with Python MNE package, and Machine Learning libraries such as Tensorflow, Keras, Sklearn are utilized to obtain CNN and further analysis.

All models are trained on an NVIDIA GeForce RTX 3060 TI GPU, with CUDA 11 and cuDNN v7, in Tensorflow [20], using the Keras API [21]. The models are

executed within Anaconda distribution platform in Jupyter Notebook environment.

3.5 Data Analysis and Performance Evaluations

3.5.1 ERP Analysis

EEG data were initially down-sampled to 100 Hz, and then epochs were obtained by extracting individual trials in the interval [-100 1000 ms] reference the stimulus start, followed by baseline-correction by subtracting mean amplitudes in the -100 to 0 ms pre-stimulus interval.

Data analysis is mainly divided into two parts. First of all, ERP responses for *NT*, *PC* and *AC* are examined in their respective configurations (*normal-speller*, and *dot-speller*). All trials from the training and test stages are concatenated in each session. After that, all of the subjects' Grand Averaged ERP patterns are observed. The difference in temporal ERP responses for target and non-target trials was statistically investigated using the signed r-squared value [44]. To evaluate the spelling performances, decoding accuracy and information transfer rates (ITRs) were calculated together with the particular sequences [45]. The classifier parameters were built using the training data, and the decoding accuracy was assessed using the test dataset.

K intervals with a step length of 20 ms and an interval length of 100 ms (e.g., $k = \{[0-100], [20-120], ...\}$) were developed during the training phase. The ERP trials across all channels were used to calculate the mean amplitude characteristics in k time intervals.

The signed r-squared value [44] was applied to statistically investigate the differences in temporal ERP responses across all channels, which calculates the point-wise separability along with the mean amplitude features k for target and non-target trials.

$$r(k) = \frac{\sqrt{N_1 \cdot N_2}}{N_1 + N_2} \frac{mean\{k|y_1\} - mean\{k|y_2\}}{std\{k\}}, \tag{3.1}$$

where k and y represent the ERP feature at a certain time interval and the class

label (i.e., target and non-target), respectively. To represent temporal-spatial features, we first selected the significant time intervals k_i based on r values across the individual channels and then concatenated them over the channel dimensions. Therefore, ERP feature vectors were formed as $\mathbf{x} \in \mathbb{R}^{D\times 1}$, where D is $ch \times k_i$. $\mathbf{X} \in \mathbb{R}^{D\times N}$ is the set of $\mathbf{x}$, where N is the total number of trials.

As our experimental systems were designed, based on a novel dot-based speller with specific mental tasks, we didn't assume any previously established data distributions of meaningful features at certain spatial (e.g., Cz, Oz) or temporal (e.g., N200, P300) locations [21, 46, 47]. Here we applied mutual information that is a non-parametric measure of relevance between binary variables $\mathbf{x}_\omega$ ($\omega = \{1, 2\}$) to find the discriminant spatial-temporal feature space. Mutual information I for the feature dimension D is then defined as:

$$I_D(\mathbf{x}; \omega) = H(\omega) - H(\omega|x), \tag{3.2}$$

where the $\mathbf{x}$ denotes feature vector and, $H(\omega)$ and $H(\mathbf{x}|\omega)$ denote entropy and conditional entropy, respectively. Entropy and conditional entropy were then defined as follows:

$$H(\omega) = -\sum_{\omega} p(\omega) \cdot log(p(\omega)), \tag{3.3}$$

$$H(\mathbf{x}|c) = -\sum_{\omega}\sum_{n=1}^{N}\int p(\omega|\mathbf{x}) log(p(\omega|\mathbf{x})). \tag{3.4}$$

Probability density functions $p(\omega)$ and $p(\omega|\mathbf{x})$ were estimated based on the Parzen window density estimation [48]. We obtained mutual information through all individual feature dimensions. Finally, discriminative features of dimension d (a subset of D) which exceeded a certain threshold of Θ were selected. The threshold parameter Θ was evaluated for individual subjects by applying 10-fold nested cross-validation to the training dataset.

From the selected feature set $\mathbf{X}_d$, a regularized linear discriminant analysis (RLDA)

[49] classifier was generated. The decision function $f(\mathbf{x}) = \mathbf{w}^{\mathrm{T}} \cdot \mathbf{x} + b$ is defined as the classification output for the input feature vector $\mathbf{x}$, where $\mathbf{w}$ is the hyper-plane for separation of binary classes and b is a bias term.

Certain target stimuli were presented 10 times (i.e., 10 sequences) in all of the tests in this investigation, which is the normal procedure in ERP studies to get a valid result by accumulative average of the current trials with the prior trials [35, 47]. As a result, decoding accuracy was determined for each sequence, ranging from one to 10. For example, the averaging of epochs across all ten sequences was used to assess the decoding accuracy of the tenth sequence. The classifier output $f(\mathbf{x}_i)$ was calculated from all of the individual letters in the speller layout, which had 36 classes (the chance level was 2.77%). The target sign was picked from the speller as the estimated letter i, where $(i = 1, ..., 36)$, which had the highest classification score. Note that in all individual trials, these decision functions were employed to offer real-time feedback for the test phase.

Chapter 4

Results

4.1 ERP Responses

The P300 components are observed during both the normal- and dot-speller trials since they are based on the oddball BCI paradigm [50]. The values of peak amplitudes are averaged within 300-400 ms (i.e., P300) at the Cz channel and are equal to 1.657 ($\pm$1.125)uV, 0.902 ($\pm$1.859)uV, 1.837 ($\pm$1.122)uV, and 1.394 ($\pm$2.639)uV for normal-passive, normal-active, dot-passive, and dot-active, respectively. The peak amplitude is a demonstration of evoked P300 components at a certain time during passive concentration (PC) in comparison to active (AC).

Moreover, Late Positive Potential (LPP) is visible approximately between 700-800 ms after the stimuli, and the mean value of peak amplitudes at the Cz are 0.463 ($\pm$1.210) uV, 1.196 ($\pm$0.989) uV, 0.476 ($\pm$0.996) uV, and 1.244 ($\pm$1.105) uV for normal-passive, normal-active, dot-passive, and dot-active, respectively. In contrast with P300 component, the LPP is evoked by the active task in both spellers (see Figure 4-1).

For non-target (NT), passive (PC) and active (AC) conditions with the null hypothesis which is equal to the maximum amplitude in two specific time intervals, a Bonferroni corrected ANOVA is utilized. Before this, data distribution in all tasks is verified by the Jarque-Bera normality test, which shows that the NT condition in the 300-400 ms in both non-oddball visual and auditory paradigms was significantly

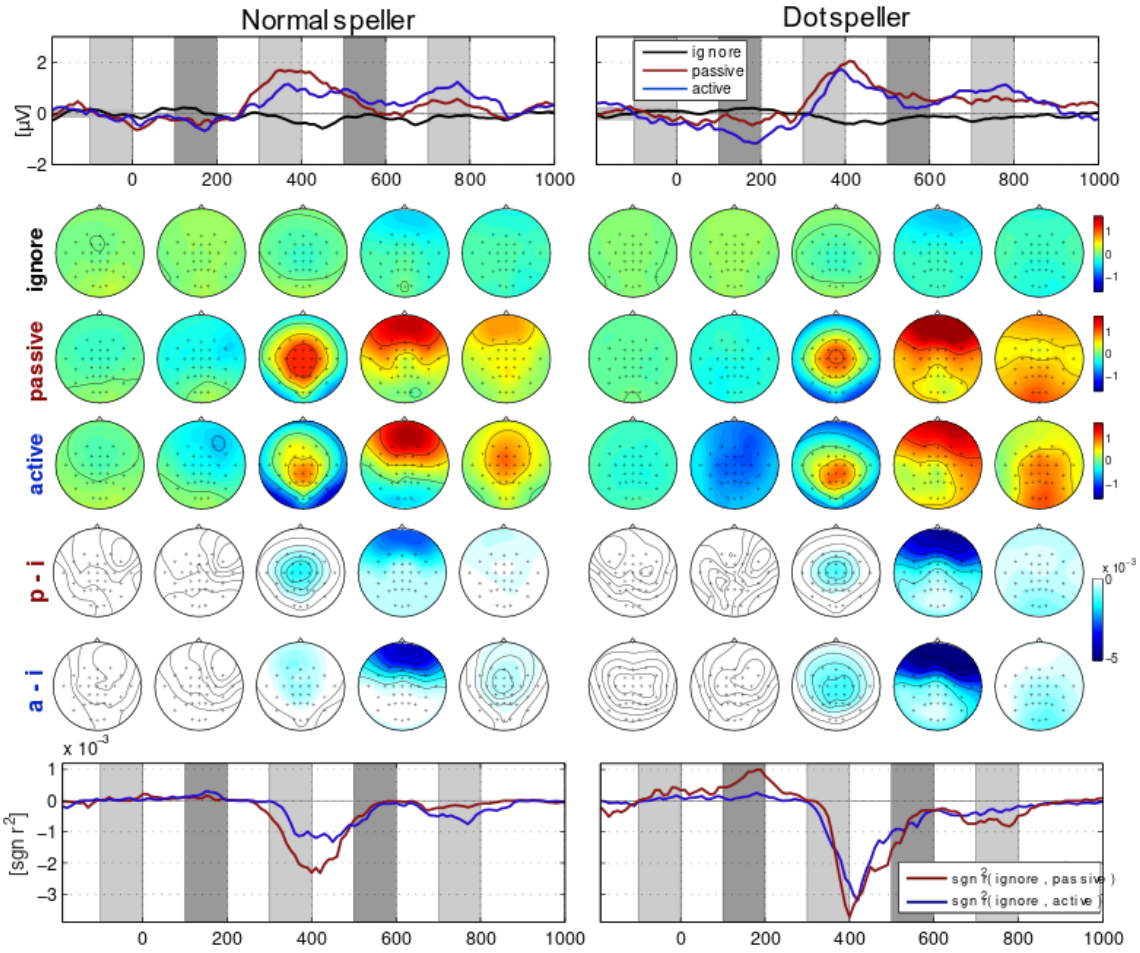

Figure (4-1) **Average ERP responses at electrode Cz.** The scalp plots demonstrate the distribution of signal response for the three different conditions, i.e., NT, PC, and AC trials. From the given ERP response graph, it can be seen that the dot-speller induces a more pronounced P300 response compared to a normal-speller. Also, passive concentration for the dot-speller induces a higher P300 amplitude than the active task.

different for the active mental task ($p < 0.05$). At the intervals of 350-450 ms and 750-800 ms, the PC and AC conditions for the spellers presented notable diversity, while for the interval of 750-800 ms in the normal-speller, the conditions were almost similar.

4.2 LDA Decoding Accuracy of BCI Speller

Figure 4-2 indicates the result of LDA classifier accuracy for normal- and dot-speller in NT vs AC and NT vs PC. Individual users' categorization accuracy, as well as average speller results, are calculated from 1 to 10 sequences (x-axis on). The results show that active tasks greatly outperformed passive activities in both speller systems. For the normal-passive, normal-active, dot-passive, and dot-active conditions, the average accuracy was 53.5%, 83.0%, 62.9%, and 88.8% after sequence four, and 76.3%, 94.1%, 78.1% and 96.8% after the ninth sequence, respectively.

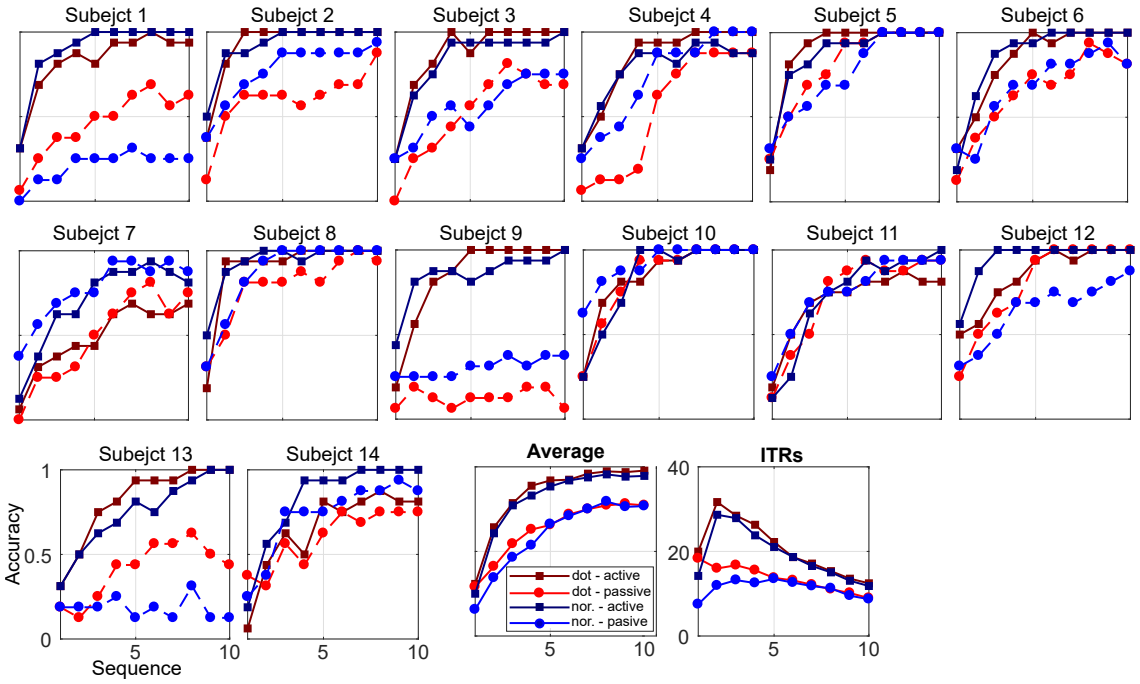

Figure (4-2) **Decoding accuracy for target and non-target discrimination in the four conditions.** The plots of decoding accuracy for 14 subjects, as well as the averaged decoding accuracy for all participants, are shown in the figure. The ITR values for each of the sequences for the four situations are shown in the last plot. Active task trials produced considerably higher decoding accuracy for each of the subjects than passive task experiments.

Active tasks ($p > 0.05$) and passive tasks ($p > 0.05$) in both speller systems show comparable patterns, according to paired t-tests. In the 2-6 and 2-8 sequences, the active tasks outperform the passive tasks for normal-speller and dot-speller, respectively. For the normal-passive, normal-active, dot-passive, and dot-active situations,

Table (4.1) **The classification accuracy on unbalanced data.** The LDA, EEGNet, SCNet=ShallowConvNet, DCNet=DeepConvNet were tested on individual sequences on data without augmentation (i.e. unbalanced dataset). The results are averaged among all 14 subjects.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
LDA	0.29	0.65	0.81	0.90	0.92	0.94	0.96	0.98	0.98	0.99
EEGNet	0.26	0.55	0.68	0.77	0.82	0.87	0.91	0.94	0.95	0.97
SCNet	0.15	0.32	0.38	0.47	0.47	0.54	0.54	0.58	0.64	**0.59**
DCNet	0.27	0.57	0.68	0.81	0.87	0.90	0.92	0.92	0.94	**0.94**

the maximum ITRs were 13.5, 28.6, 18.3, and 31.6 [bits/min], respectively.

4.3 CNN Performance

The learning capacity of CNNs can be demonstrated through the ability of automatic feature extraction from given raw signals. However, the features are not hand-crafted, and it is difficult to understand them and produce interpretable models. This is a common issue when CNNs are utilized for EEG data analysis where signal features contain noisy artifacts.

The EEGNet model is a small CNN architecture that may be employed in a wide range of EEG-based BCI devices and applications. It can also be trained with a small dataset and provide features that are neurophysiologically interpretable.

The results of trained and tested LDA and CNN models are presented in the Table 4.1 where the accuracy is calculated in individual sequences (total 10 sequence) for NT vs AC tasks, and the results were averaged across all 14 subjects. The data presented in the Table 4.1 is unbalanced dataset, whereas in Table 4.2 the augmentation technique, such as balancing tagret and non-target (i.e. the ratio of NT vs AC is 1:1). Note that the augmentation was not applied to LDA due to already satisfactory results. From both tables, it is clearly seen that the accuracy of ShallowConvNet is greatly improved after augmentation.

The computational time with accuracy at each epoch and sequence are calculated and presented in the Table 4.3 and Table 4.4 for individual and single accuracy respectively performances for EEGNet, DeepConvNet and ShallowConvNet.

Table (4.2) **The classification accuracy on balanced data.** The LDA, EEGNet, SCNet=ShallowConvNet, DCNet=DeepConvNet were tested on individual sequences on data with augmentation (i.e. balanced dataset). The results are averaged among all 14 subjects.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
LDA	0.29	0.65	0.81	0.90	0.92	0.94	0.96	0.98	0.98	0.99
EEGNet	0.30	0.64	0.78	0.86	0.87	0.93	0.96	0.96	0.97	0.97
SCNet	0.27	0.58	0.67	0.70	0.75	0.79	0.82	0.85	0.86	**0.88**
DCNet	0.33	0.63	0.78	0.83	0.89	0.94	0.95	0.96	0.96	**0.97**

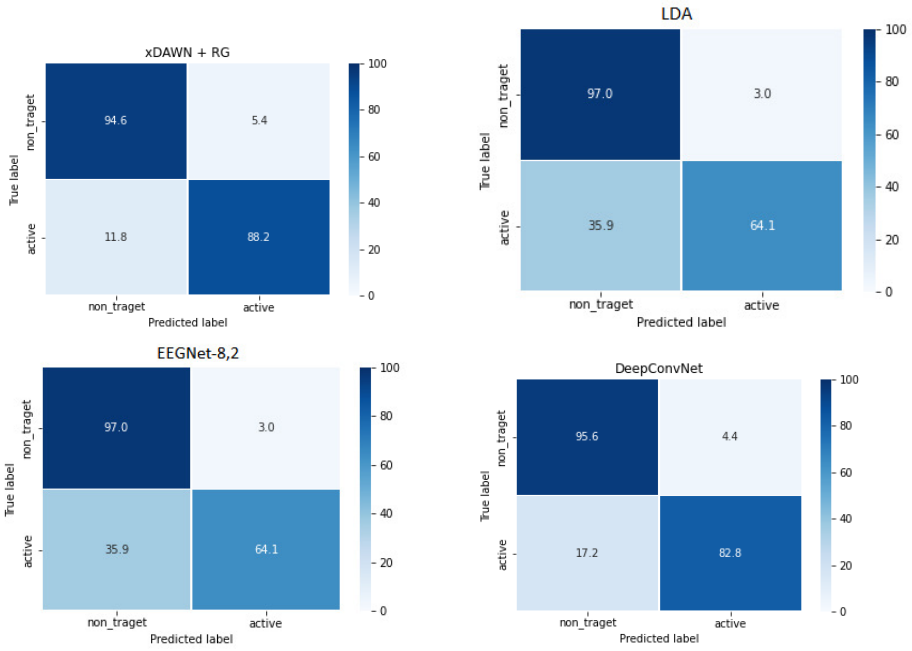

Figure (4-3) **Confusion matrix of models for NT vs. AC classes in single trial accuracy for one subject.** The figure indicates the an averaged classification accuracy for DeepConvNet, LDA, ShallowConvNet and xDAWN+RG models trained and tested on all subjects for non-target and active, and non-target and passive conditions.

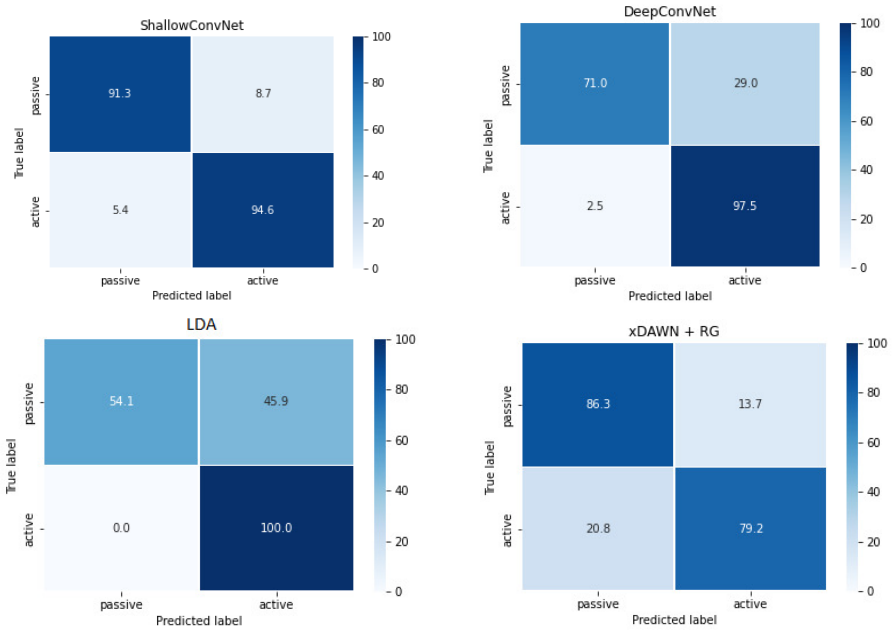

Figure (4-4) **Confusion matrix of models for PC vs. AC classes in single trial accuracy for one subject.** The figure indicates the an averaged classification accuracy for DeepConvNet, LDA, EEGNet and xDAWN+RG models trained and tested on all subjects for passive and active conditions.

Table (4.3) **The performance of CNN models with computational time in individual sequences.** The accuracy presented in % format, and the computational time is in milliseconds at each sequence of maximum 10 for EEGNet, DeepConvNet, and ShallowConvNet.

Sequence	EEGNet		DeepConvNet		ShallowConvNet	
	Accuracy	Time (ms)	Accuracy	Time (ms)	Accuracy	Time (ms)
1	29.91%	149	32.59%	170	27.23%	130
2	63.84%	138	62.95%	169	58.04%	132
3	77.68%	127	77.68%	160	66.96%	122
4	85.71%	134	83.48%	130	70.09%	133
5	87.05%	124	89.29%	150	75.00%	140
6	92.86%	159	94.20%	167	78.57%	110
7	96.43%	140	95.09%	120	82.14%	112
8	95.54%	120	95.98%	110	84.82%	109
9	97.32%	129	95.98%	112	86.16%	90
10	97.32%	110	97.32%	150	87.50%	120

Table (4.4) **The performance of CNN models with computational time in single trial.** The accuracy presented in % format, and the computational time is in milliseconds at each epoch of maximum 30 for EEGNet, DeepConvNet, and ShallowConvNet. The epochs of 30 is chosen as the most effective number of epochs.

Epoch	EEGNet		DeepConvNet		ShallowConvNet	
	Acc (%)	Time (ms)	Acc (%)	Time (ms)	Acc (%)	Time (ms)
1	79.17%	220	70.11%	232	64.22%	190
2	70.61%	221	71.11%	145	79.01%	193
3	71.66%	225	71.11%	225	68.12%	167
4	76.64%	236	80.22%	211	75.62%	200
5	76.64%	243	80.99%	243	76.64%	190
6	76.64%	163	81.78%	123	76.64%	163
7	80.56%	108	82.90%	111	80.56%	108
8	81.56%	209	81.56%	97	81.56%	209
9	82.57%	193	82.57%	193	81.21%	193
10	86.06%	188	85.67%	188	86.06%	122
11	80.56%	211	86.78%	109	80.56%	211
12	90.56%	196	90.56%	300	90.56%	176
13	93.06%	158	92.11%	158	93.06%	186
14	93.06%	191	93.06%	191	93.06%	191
15	93.06%	185	93.06%	185	92.11%	185
16	93.06%	152	93.06%	152	93.06%	152
17	94.44%	185	96.88%	185	94.44%	185
18	94.44%	152	94.44%	124	95.42%	178
19	94.44%	198	96.77%	198	94.44%	198
20	95.14%	204	95.14%	204	95.14%	199
21	95.14%	195	92.33%	195	97.91%	299
22	95.14%	201	98.99%	201	95.14%	201
23	95.14%	194	95.14%	194	95.14%	194
24	96.53%	172	96.53%	172	96.53%	172
25	96.53%	203	96.53%	203	96.53%	300
26	96.53%	150	97.88%	122	96.53%	277
27	97.22%	188	97.22%	188	97.22%	188
28	97.22%	195	97.22%	111	90.12%	195
29	97.92%	120	95.12%	120	97.92%	108
30	99.31%	130	97.22%	140	89.11%	99

4.4 Discussion

The performance of a speller system that blends pixel-level visual stimuli with sound imagery stimuli in order to evaluate the practical viewpoint of real-world BCI applications (dot-speller) was observed. The proposed speller systems rely on oddball paradigm performance, but with a focus on a less obtrusive visual interface. To test the efficacy of tiny visual stimulation, we compared the results of a traditional speller that flashed the letters themselves to our innovative dot-speller that flashed 0.1 mm dots instead of the letters on a smartphone. We also looked at how useful active sound imagery activities were in this scenario versus passive staring. Our findings have potentially significant implications for future BCI interfaces.

There were two significant intervals for discriminating AC and PC when it came to purposeful command: 300-400 ms (P300) and 700-800 ms (LPP) (see Figure 4-1). Interestingly, the PC task had a larger P300 component than the AC task, whereas the LPP had the reverse outcome. We believe that the LPP is derived from the user's active mental job and that it is a more potent feature than the P300 component. While the P300 response is an external response to the unusual stimulus [51], passive attention can lead to incorrect orders, whereas active mental attention is far more trustworthy in terms of purpose.

Figure 4-2 shows that the active task greatly exceeded the passive attention task in terms of intention. In the normal- and dot-layout systems, the average spelling accuracy for the passive job was 76.3 percent and 78.1 percent, respectively. This accuracy is lower than in prior research, which showed more than 90% accuracy [21, 46, 52, 53].

The speller system is implemented on a smartphone, which accounts for the lesser performance. For the normal-speller letters, the suggestive stimulus sizes were less than 0.8 cm, and for the dot-speller pattern, they were barely 0.1 mm. As previously discovered, this occurs because smaller-sized and closer-positioned letters limit ERP responses, lowering system performance, according to [54]. Despite this flaw, the active tasks' spelling performance was 94.1 percent and 96.8 percent, respectively.

This outcome is on par with, if not better than, many advanced spelling systems [46, 22, 52, 53, 55].

In individual sequences LDA model outperformed CNNs (see Table 4.1 and Table 4.2), and the accuracy were 99%, 97%,88% and 97% for LDA, EEGNet, ShallowConvNet, and DeepConvNet, respectively. It is difficult to beat LDA in sequential classification because one of the advantages of LDA is that it uses information from both the features to create a new axis which in turn minimizes the variance and maximizes the class distance of the two variables. CNN can be only comparable and close to LDA, or the data augmentation, such as balancing data should be applied before feeding it to CNN to achieve better results. However, in single trial classification all CNNs demonstrated better performance than LDA. This might be caused by the fact that the input for LDA was not averaged among sample points (i.e. from 100 to 10), so the data was taken as it is (i.e. input shape ((14,32,101) where 14-subject number, 32-channels, 101-sample points). In other words, due to high dimensional data (i.e input size of time sample was 100, in comparison to sequential where we took 10 time sample points) the LDA classification leads to overfitting.

Moreover, during experiments two conditions PC and AC were analyzed in all models. According to Figure 4-4 the classifiers can distinguish between passively gazing and actively choosing target character. This means that future BCI spellers can utilize such opportunity to show intentional and unintentional gazing. LDA classifier has one of the lowest accuracy due to that it requires explicitly explained categories (i.e. classes), and PC and AC are similar (i.e. they both are target classes, and are meant to show the user intention to choose the desired output character).

4.5 Limitation of the Study

In BCI experiments there are several limitations that can be divided into three categories, such as small sample size, usability or practicability and duration. The first one means that there are the limited number of volunteers for conducting experiments, so it is difficult to find users for BCI spellers. The second one is that in this

study the experiments were conducted on healthy people with normal or corrected vision. Thirdly, the duration might be inconvenient to users, and it is important to allocate enough time for calibration, training and testing, and also for BCI headset and electrodes installments.

In classification experiments, there are also limitations, such as dataset suitability and subject dependency. First one means that the open-sourced models were tested on different data with different experiments, so applying and fitting own dataset is time consuming and gives different results. Secondly, the amplitude and latency of P300 wave are subject dependent. In other words, the wave amplitude has a positive association with the uncertainty of occurrence of the target stimuli while its latency depends on the time needed for a subject to cognitively evaluate the target stimuli.

Chapter 5

Conclusion and Future Work

In this research the proposed BCI system is tested in real-world condition which makes it practical to use for further investigation. The concept is maintaining high classification accuracy by generating endogenous ERP components which are activated when the user performs a mental task (sound imagery). Minor subjects participated in the experiments are healthy, with normal or corrected vision, and therefore, further research should be conducted with different condition, such as low vision or other eye impairments. This research contributes to the development of a practical and efficient BCI speller, which includes a simple headset design, a powerful CNN-based classifier, and a new spelling pattern. The speller system is implemented on a smartphone with stimulus size for normal-speller is 0.8 cm, and for dot-speller is 0.1 mm. It is mentioned that smaller stimuli size leads to the lower performance. In other words, smaller-sized and closer-positioned letters limit ERP responses which reduces system performance. Despite this flaw, the dot-speller demonstrated comparable results for all models. It was achieved by generating endogenous ERP components by introducing simple mental task. Moreover, Passive Concentration (simple gazing) and Active Concentration (mental task) on normal-speller and dot-speller were compared, and in the result, mental exercise considerably increased performance, and the dot-speller demonstrated high accuracy. Average spelling accuracy for LDA in individual sequences are following: passive tasks for normal -layout 76.3% and for the proposed dot-layout - 78.1%, while active tasks for normal - 94.1% and dot-layouts - 96.8%.

Average spelling accuracy for CNN results in individual sequences for non-target and active tasks for dot-speller are the following: before augmenting the data EEGNet - 97%, ShallowConvNet - 59%, DeepConvNet - 94%, and after augmentation EEGNet - 97%, ShallowConvNet - 88%, DeepConvNet - 97%.

The experiments were done on healthy (normal or corrected vision) people, so further research is needed to determine whether the suggested feature is suitable for users with vision impairments. Moreover, the new spelling layout presented in this paper has the potential to make future BCI spelling systems more reliable and user-friendly. Additionally, preliminary results of passive versus active tasks classification can be used to develop new BCI speller with intentionally and unintentionally gazing. Finally, for further classification experiments predictive variables (e.g. age, gender, BCI experience) might be included for better performance.

Appendix A

Tables

A.1 Unbalanced dataset

The Tables below demonstrate the models performances in individual sequences for all subjects (rows) for 10 sequences (columns) on unbalanced data before augmentation.

LDA

0.3125	0.8125	0.875	0.9375	0.9375	1	1	1	1	1
0.5	0.8125	0.75	0.9375	0.9375	1	1	1	1	1
0.3125	0.6875	0.9375	1	0.9375	0.9375	0.9375	0.9375	1	1
0.1875	0.5	0.75	0.9375	0.9375	0.9375	0.9375	0.9375	0.875	0.9375
0.25	0.625	0.75	0.875	1	1	1	1	1	1
0.25	0.5625	0.875	0.875	0.9375	0.9375	1	1	1	1
0.1875	0.3125	0.6875	0.6875	0.6875	0.6875	0.75	0.875	0.8125	0.9375
0.4375	0.8125	0.875	0.875	0.9375	0.9375	1	1	1	1
0.375	0.75	0.75	0.875	0.875	0.9375	0.9375	1	1	1
0.3125	0.5625	0.875	1	1	1	1	1	1	1
0.125	0.625	0.8125	0.875	0.875	0.875	1	1	1	1
0.5625	0.875	1	1	1	1	1	1	1	1
0.1875	0.5625	0.5625	0.8125	0.875	0.875	0.9375	1	1	1
0.0625	0.5625	0.875	0.9375	1	1	1	1	1	1
0.2901785714	**0.6473214286**	**0.8125**	**0.9017857143**	**0.9241071429**	**0.9375**	**0.9642857143**	**0.9821428571**	**0.9776785714**	**0.9910714286**

EEGNET

0.25	0.8125	0.875	0.9375	0.9375	0.9375	1	1	1	1
0.375	0.75	0.75	0.9375	1	1	1	1	1	1
0.125	0.625	0.625	0.8125	0.875	0.9375	0.9375	1	1	1
0.125	0.25	0.3125	0.5	0.4375	0.5	0.5625	0.625	0.75	0.75
0.25	0.5	0.625	0.6875	0.6875	0.875	0.9375	1	1	1
0.25	0.375	0.8125	0.6875	0.875	1	1	1	1	1
0.0625	0.25	0.5625	0.5625	0.625	0.75	0.8125	0.75	0.625	0.8125
0.4375	0.8125	0.8125	0.875	0.875	0.875	0.9375	1	1	1
0.375	0.8125	0.8125	0.875	0.875	0.875	0.875	0.875	0.875	1
0.3125	0.5625	0.6875	0.75	0.875	0.875	1	1	1	1
0.125	0.4375	0.75	0.875	0.875	0.9375	1	1	1	1
0.4375	0.8125	0.9375	1	1	1	1	1	1	1
0.25	0.3125	0.4375	0.625	0.75	0.8125	0.8125	0.875	1	1
0.25	0.4375	0.5	0.625	0.8125	0.8125	0.875	1	1	1
0.2589285714	**0.5535714286**	**0.6785714286**	**0.7678571429**	**0.8214285714**	**0.8705357143**	**0.9107142857**	**0.9375**	**0.9464285714**	**0.96875**

SCNET

0.25	0.6875	0.8125	0.8125	0.8125	0.875	0.9375	0.9375	1	1
0.3125	0.5625	0.8125	0.9375	0.9375	1	1	1	1	1
0.125	0.3125	0.125	0.4375	0.375	0.3125	0.4375	0.4375	0.5	0.4375
0.0625	0.0625	0.3125	0.1875	0.0625	0.1875	0.1875	0.25	0.375	0.375
0.1875	0.125	0.125	0.3125	0.375	0.5625	0.4375	0.5	0.5625	0.5625
0.125	0.25	0.25	0.375	0.375	0.4375	0.5625	0.5625	0.625	0.5
0	0.125	0.125	0.125	0.1875	0.0625	0.0625	0.0625	0.125	0.0625
0	0.25	0.375	0.5625	0.625	0.5625	0.5625	0.5625	0.625	0.5625
0.1875	0.6875	0.625	0.8125	0.875	0.875	0.875	0.9375	0.9375	0.875
0.125	0.3125	0.5	0.5	0.4375	0.5625	0.375	0.375	0.5	0.375
0.0625	0.125	0.1875	0.1875	0.125	0.1875	0.125	0.125	0.3125	0.125
0.3125	0.625	0.5625	0.8125	0.75	0.875	0.875	0.875	0.9375	0.9375
0.1875	0.25	0.25	0.25	0.375	0.625	0.625	0.8125	0.75	0.75
0.1875	0.0625	0.25	0.3125	0.3125	0.4375	0.5625	0.625	0.6875	0.75
0.1517857143	**0.3169642857**	**0.3794642857**	**0.4732142857**	**0.4732142857**	**0.5401785714**	**0.5446428571**	**0.5758928571**	**0.6383928571**	**0.59375**

DCNET

0.3125	0.75	0.8125	0.875	0.9375	0.9375	1	1	1	1
0.3125	0.6875	0.9375	1	1	1	1	1	1	1
0.3125	0.5	0.625	0.6875	0.875	0.9375	1	1	1	1
0.1875	0.4375	0.5	0.6875	0.6875	0.6875	0.6875	0.6875	0.75	0.8125
0.125	0.5625	0.6875	0.9375	0.8125	0.9375	0.9375	1	1	1
0.3125	0.5	0.6875	0.8125	0.9375	0.9375	0.9375	0.8125	0.8125	0.9375
0.1875	0.4375	0.5625	0.625	0.625	0.6875	0.75	0.6875	0.75	0.625
0.5625	0.8125	0.8125	0.9375	1	1	0.9375	1	0.9375	1
0.375	0.75	0.875	0.9375	0.875	0.875	0.9375	0.9375	1	1
0.125	0.3125	0.6875	0.875	0.9375	1	1	0.9375	1	1
0.125	0.4375	0.625	0.75	0.75	0.8125	0.9375	0.9375	1	0.9375
0.375	0.9375	0.875	1	1	1	1	1	1	1
0.1875	0.3125	0.1875	0.375	0.8125	0.8125	0.8125	0.8125	0.875	0.875
0.3125	0.5625	0.625	0.875	0.875	0.9375	0.9375	1	1	1
0.2723214286	**0.5714285714**	**0.6785714286**	**0.8125**	**0.8660714286**	**0.8973214286**	**0.9196428571**	**0.9151785714**	**0.9375**	**0.9419642857**

A.2 Balanced dataset

The Tables below demonstrate the models performances in individual sequences for all subjects (rows) for 10 sequences (columns) on balanced data after augmentation.

LDA

0.3125	0.8125	0.875	0.9375	0.9375	1	1	1	1	1
0.5	0.8125	0.75	0.9375	0.9375	1	1	1	1	1
0.3125	0.6875	0.9375	1	0.9375	0.9375	0.9375	0.9375	1	1
0.1875	0.5	0.75	0.9375	0.9375	0.9375	0.9375	0.9375	0.875	0.9375
0.25	0.625	0.75	0.875	1	1	1	1	1	1
0.25	0.5625	0.875	0.875	0.9375	0.9375	1	1	1	1
0.1875	0.3125	0.6875	0.6875	0.6875	0.6875	0.75	0.875	0.8125	0.9375
0.4375	0.8125	0.875	0.875	0.9375	0.9375	1	1	1	1
0.375	0.75	0.75	0.875	0.875	0.9375	0.9375	1	1	1
0.3125	0.5625	0.875	1	1	1	1	1	1	1
0.125	0.625	0.8125	0.875	0.875	0.875	1	1	1	1
0.5625	0.875	1	1	1	1	1	1	1	1
0.1875	0.5625	0.5625	0.8125	0.875	0.875	0.9375	1	1	1
0.0625	0.5625	0.875	0.9375	1	1	1	1	1	1
0.2901785714	**0.6473214286**	**0.8125**	**0.9017857143**	**0.9241071429**	**0.9375**	**0.9642857143**	**0.9821428571**	**0.9776785714**	**0.9910714286**

EEGNET

0.3125	0.6875	0.875	0.9375	0.875	0.9375	1	1	1	1
0.4375	0.75	0.875	0.875	0.9375	1	1	1	1	1
0.25	0.5	0.75	0.875	0.8125	0.875	0.9375	0.9375	0.9375	1
0.3125	0.4375	0.75	0.8125	0.875	0.875	0.875	0.875	0.875	0.875
0.125	0.625	0.75	0.8125	0.875	1	1	1	1	1
0.3125	0.6875	0.6875	0.875	0.875	1	1	1	1	1
0.0625	0.1875	0.5	0.625	0.5625	0.75	0.875	0.8125	0.8125	0.8125
0.5	0.8125	0.875	1	1	1	1	1	1	1
0.3125	0.8125	0.8125	0.9375	0.9375	0.875	1	1	1	1
0.3125	0.875	1	0.9375	1	1	1	1	1	1
0.25	0.5	0.8125	0.875	0.9375	0.9375	1	1	1	0.9375
0.5	0.875	1	1	1	1	1	1	1	1
0.125	0.5	0.3125	0.5	0.5625	0.75	0.8125	0.75	1	1
0.375	0.6875	0.875	0.9375	0.9375	1	1	1	1	1
0.2991071429	**0.6383928571**	**0.7767857143**	**0.8571428571**	**0.8705357143**	**0.9285714286**	**0.9642857143**	**0.9553571429**	**0.9732142857**	**0.9732142857**

SCNET

0.1875	0.6875	0.875	0.9375	0.875	1	1	1	1	1
0.3125	0.75	0.9375	0.9375	1	1	1	1	1	1
0.25	0.5625	0.8125	0.875	1	1	1	0.9375	0.9375	0.9375
0.125	0.25	0.375	0.625	0.5	0.6875	0.875	0.8125	0.8125	0.6875
0.1875	0.5	0.5	0.75	0.875	0.9375	0.9375	0.9375	1	1
0.125	0.1875	0.625	0.6875	0.6875	0.625	0.8125	0.875	0.875	0.9375
0.1875	0.25	0.5	0.5	0.4375	0.5625	0.6875	0.6875	0.75	0.6875
0.4375	0.875	0.8125	0.8125	0.9375	1	1	1	1	1
0.1875	0.75	0.875	0.8125	0.75	0.8125	0.875	1	1	1
0.1875	0.375	0.75	0.8125	0.9375	0.9375	1	1	0.9375	1
0.125	0.3125	0.6875	0.6875	0.8125	0.75	0.875	0.9375	0.9375	0.875
0.4375	0.6875	0.6875	0.75	0.875	0.875	1	1	1	1
0.1875	0.3125	0.5	0.5	0.625	0.8125	0.8125	0.875	0.9375	0.8125
0.1875	0.25	0.5625	0.75	0.875	1	1	1	1	1
0.2232142857	**0.4821428571**	**0.6785714286**	**0.7455357143**	**0.7991071429**	**0.8571428571**	**0.9196428571**	**0.9330357143**	**0.9419642857**	**0.9241071429**

DCNET

0.3125	0.75	0.875	1	1	0.9375	0.9375	1	1	1
0.3125	0.8125	0.9375	0.9375	1	1	1	1	1	1
0.125	0.5625	0.5625	0.8125	0.9375	0.875	0.9375	1	1	1
0.1875	0.6875	0.6875	0.8125	0.8125	0.75	0.8125	0.875	0.75	0.8125
0.3125	0.5625	0.6875	0.8125	0.9375	0.875	0.9375	1	1	1
0.1875	0.0625	0	0.0625	0.125	0.3125	0.25	0.25	0.3125	0.4375
0.1875	0.3125	0.625	0.5625	0.5625	0.6875	0.8125	0.8125	0.6875	0.8125
0.4375	0.8125	0.75	0.875	0.9375	1	0.9375	1	1	1
0.375	0.75	0.8125	0.875	0.875	0.875	0.875	0.875	0.9375	0.9375
0.1875	0.375	0.625	0.625	0.625	0.625	0.75	0.6875	0.6875	0.625
0.125	0.5625	0.75	0.9375	0.9375	1	1	1	1	0.9375
0.4375	0.8125	0.8125	1	1	1	1	1	1	1
0.1875	0.25	0.25	0.6875	0.6875	0.8125	0.75	0.8125	0.8125	0.9375
0.1875	0.3125	0.5	0.6875	0.8125	0.875	0.75	0.875	0.75	0.875
0.2544642857	**0.5446428571**	**0.6339285714**	**0.7633928571**	**0.8035714286**	**0.8303571429**	**0.8392857143**	**0.8705357143**	**0.8526785714**	**0.8839285714**

Appendix A

Figures

The confusion matrices below indicate three class classification trained and tested on ShallowConvNet model: NT vs T (AC and PC) which is unbalanced dataset. So NT sample is larger that AC and PC which leads to misclassification. So to increase performance we can apply augmentation technique.

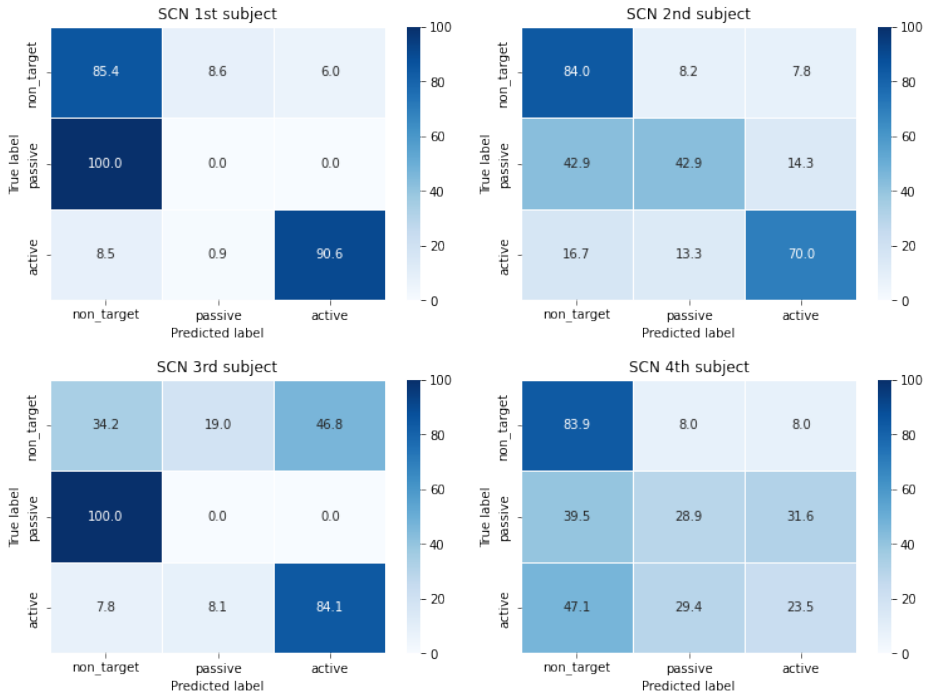

Figure (A-1) **Confusion matrix of models for three classes NT, PC and AC in single trial accuracy for subject 1, 2, 3 and 4, respectively.**

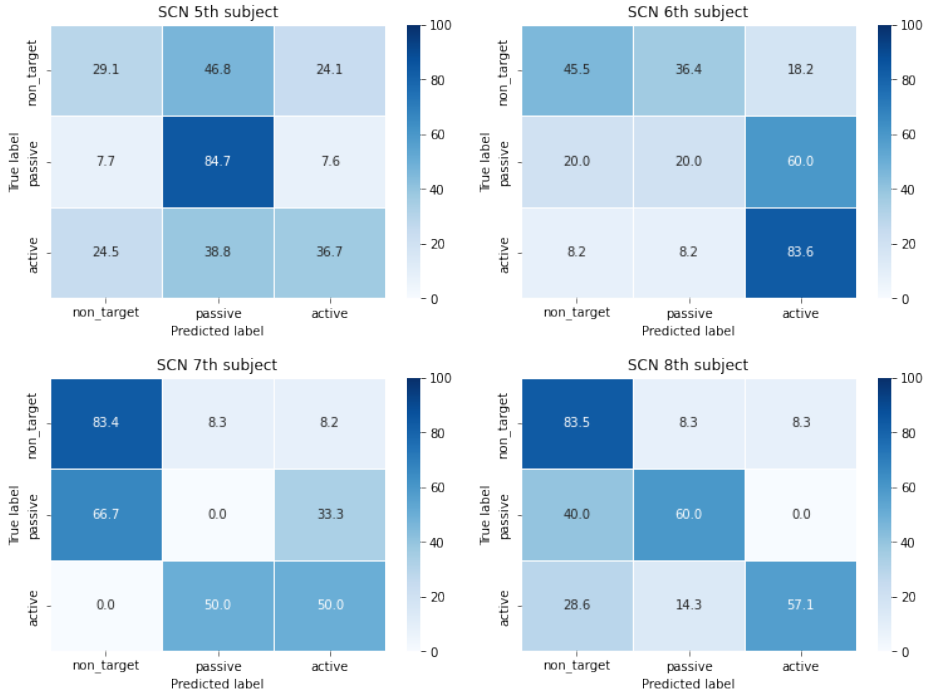

Figure (A-2) **Confusion matrix of models for three classes NT, PC and AC in single trial accuracy for subject 5, 6, 7, and 8, respectively.**

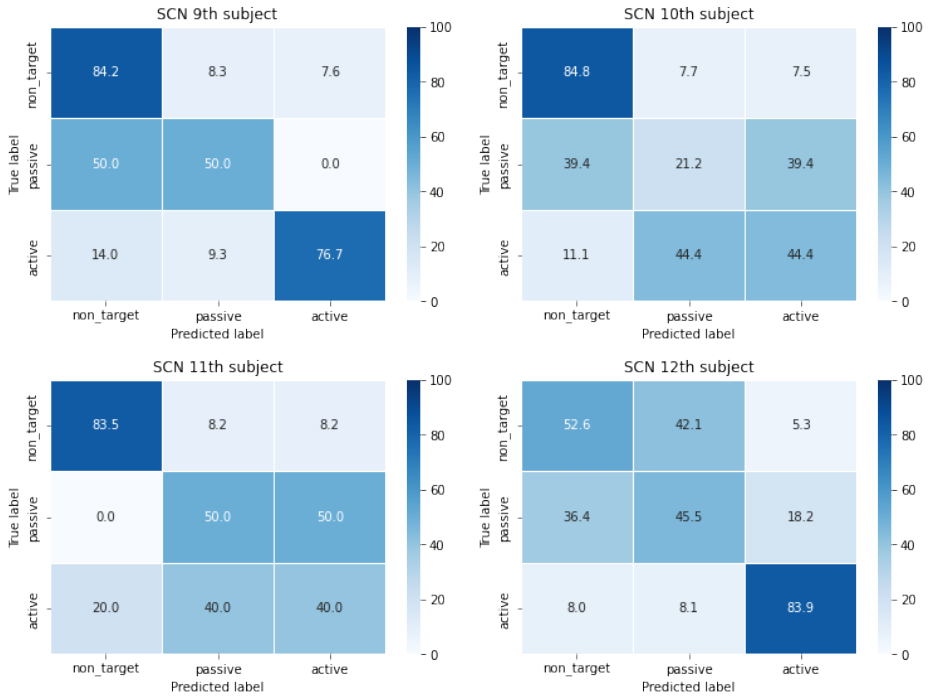

Figure (A-3) **Confusion matrix of models for three classes NT, PC and AC in single trial accuracy for subject 9, 10, 11, and 12, respectively.**

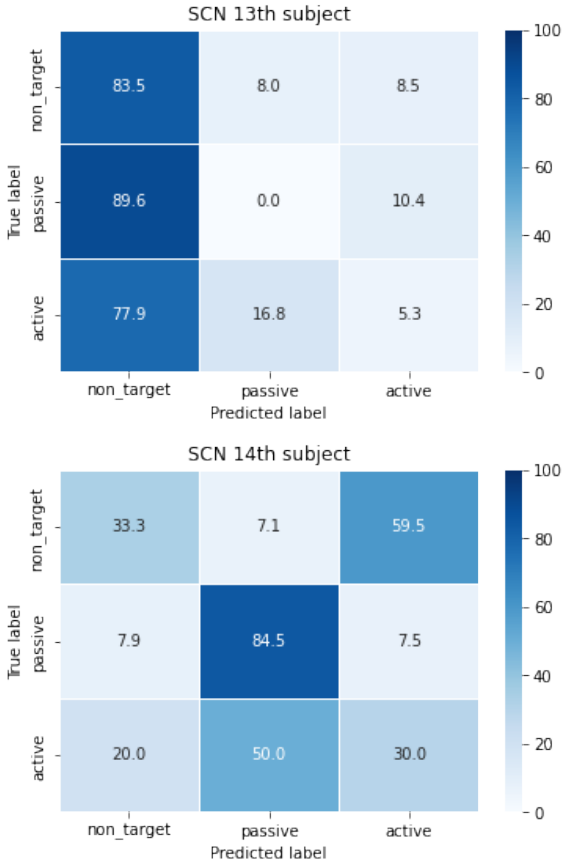

Figure (A-4) **Confusion matrix of models for three classes NT, PC and AC in single trial accuracy for subject 13 and 14, respectively.**

Bibliography

[1] Jing Jin, Zongmei Chen, Ren Xu, Yangyang Miao, Xing yu Wang, and Tzyy-Ping Jung. Developing a novel tactile P300 brain-computer interface with a cheeks-stim paradigm. *IEEE Transactions on Biomedical Engineering*, 2020.

[2] Gomez-Gil J. Nicolas-Alonso, L.F. Brain-computer interfaces, a review. *Sensors Journal*, 12(2):1211–1212, 2012.

[3] Milan Brázdil, Michal Mikl, Radek Mareček, Petr Krupa, and Ivan Rektor. Effective connectivity in target stimulus processing: a dynamic causal modeling study of visual oddball task. *Neuroimage*, 35(2):827–835, 2007.

[4] Miseon Shim, Do-Won Kim, Seung-Hwan Lee, and Chang-Hwan Im. Disruptions in small-world cortical functional connectivity network during an auditory oddball paradigm task in patients with schizophrenia. *Schizophrenia research*, 156(2-3):197–203, 2014.

[5] Kwon O.Y.-Kim Y.J. Kim H.K. Lee Y.E. Williamson J. Fazli S. Lee S.W. Lee, M.H. Eeg dataset and openbmi toolbox for three bci paradigms: an investigation into bci illiteracy. *GigaScience Journal*, 8(5), 2019.

[6] Jeong Woo Choi, Ki-Young Jung, Chi Hyun Kim, and Kyung Hwan Kim. Changes in gamma-and theta-band phase synchronization patterns due to the difficulty of auditory oddball task. *Neuroscience letters*, 468(2):156–160, 2010.

[7] Luis Fernando Nicolas-Alonso and Jaime Gomez-Gil. Brain-computer interfaces, a review. *Sensors*, 12(2):1211–1279, 2012.

[8] Albina Li, Kanat Alimanov, Siamac Fazli, and Min-Ho Lee. Towards paradigm-independent brain computer interfaces. In *2020 8th International Winter Conference on Brain-Computer Interface (BCI)*, pages 1–6. IEEE, 2020.

[9] Nader Karamzadeh, Andrei Medvedev, Afrouz Azari, Amir Gandjbakhche, and Laleh Najafizadeh. Capturing dynamic patterns of task-based functional connectivity with EEG. *NeuroImage*, 66:311–317, 2013.

[10] Youguo Chen, Xiting Huang, Yangmei Luo, Chunhua Peng, and Chunxiang Liu. Differences in the neural basis of automatic auditory and visual time perception: ERP evidence from an across-modal delayed response oddball task. *Brain research*, 1325:100–111, 2010.

[11] JD Kropotov, Risto Näätänen, AV Sevostianov, Kimmo Alho, K Reinikainen, and OV Kropotova. Mismatch negativity to auditory stimulus change recorded directly from the human temporal cortex. *Psychophysiology*, 32(4):418–422, 1995.

[12] Akshansh Gupta, Ramesh Kumar Agrawal, Jyoti Singh Kirar, Javier Andreu-Perez, Wei-Ping Ding, Chin-Teng Lin, and Mukesh Prasad. On the utility of power spectral techniques with feature selection techniques for effective mental task classification in noninvasive BCI. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2019.

[13] Winko W An, Alexander Pei, Abigail L Noyce, and Barbara Shinn-Cunningham. Decoding auditory attention from single-trial eeg for a high-efficiency brain-computer interface. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 3456–3459. IEEE, 2020.

[14] Kensho Hara, Hirokatsu Kataoka, and Yutaka Satoh. Learning spatio-temporal features with 3d residual networks for action recognition. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 3154–3160, 2017.

[15] Yanfei Kang, Rob J Hyndman, and Feng Li. GRATIS: Generating time series with diverse and controllable characteristics. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 13(4):354–376, 2020.

[16] No-Sang Kwak, Klaus-Robert Müller, and Seong-Whan Lee. A convolutional neural network for steady state visual evoked potential classification under ambulatory environment. *PloS one*, 12(2):e0172578, 2017.

[17] Boris Kleber and Niels Birbaumer. Direct brain communication: neuroelectric and metabolic approaches at tübingen. *Cognitive Processing*, 6(1):65–74, 2005.

[18] Hartmut Heinrich, Holger Gevensleben, Franz Joseph Freisleder, Gunther H Moll, and Aribert Rothenberger. Training of slow cortical potentials in attention-deficit/hyperactivity disorder: evidence for positive behavioral and neurophysiological effects. *Biological psychiatry*, 55(7):772–775, 2004.

[19] Erwei Yin, Zongtan Zhou, Jun Jiang, Fanglin Chen, Yadong Liu, and Dewen Hu. A novel hybrid BCI speller based on the incorporation of SSVEP into the P300 paradigm. *Journal of Neural Engineering*, 10(2):026012, 2013.

[20] L.A. Farwell and E. Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 70(6):510–523, 1988.

[21] Seul-Ki Yeom, Siamac Fazli, Klaus-Robert Müller, and Seong-Whan Lee. An efficient ERP-based brain-computer interface using random set presentation and face familiarity. *PloS one*, 9(11), 2014.

[22] Qi Li, Zhaohua Lu, Ning Gao, and Jingjingc Yang. Optimizing the performance of the visual P300-speller through active mental tasks based on color distinction and modulation of task difficulty. *Frontiers in human neuroscience*, 13:130, 2019.

[23] Min-Ho Lee, John Williamson, Young-Jin Kee, Siamac Fazli, and Seong-Whan Lee. Robust detection of event-related potentials in a user-voluntary short-term imagery task. *PloS one*, 14(12):e0226236, 2019.

[24] Min-Ho Lee, John Williamson, Young-Eun Lee, and Seong-Whan Lee. Mental fatigue in central-field and peripheral-field steady-state visually evoked potential and its effects on event related potential responses. *NeuroReport*, 29(15), 2018.

[25] Z Kubová, J Kremlacek, J Szanyi, J Chlubnová, and M Kuba. Visual event-related potentials to moving stimuli: normative data. *Physiological Research*, 51(2):199–204, 2002.

[26] Thilo Hinterberger, Stefan Schmidt, Nicola Neumann, Jürgen Mellinger, Benjamin Blankertz, Gabriel Curio, and Niels Birbaumer. Brain-computer communication and slow cortical potentials. *IEEE Transactions on Biomedical Engineering*, 51(6):1011–1018, 2004.

[27] Xiaolin Xiao, Minpeng Xu, Jing Jin, Yijun Wang, Tzyy-Ping Jung, and Dong Ming. Discriminative canonical pattern matching for single-trial classification of ERP components. *IEEE Transactions on Biomedical Engineering*, 2019.

[28] Minpeng Xu, Xiaolin Xiao, Yijun Wang, Hongzhi Qi, Tzyy-Ping Jung, and Dong Ming. A brain–computer interface based on miniature-event-related potentials induced by very small lateral visual stimuli. *IEEE Transactions on Biomedical Engineering*, 65(5):1166–1175, 2018.

[29] John Polich, Patricia Crane Ellerson, and Jill Cohen. P300, stimulus intensity, modality, and probability. *International Journal of Psychophysiology*, 23(1):55–62, 1996.

[30] Zhimin Lin, Chi Zhang, Ying Zeng, li Tong, and Bin Yan. A novel p300 bci speller based on the triple rsvp paradigm. *Scientific Reports*, 8, 02 2018.

[31] Maarten de Vos, Markus Kroesen, Reiner Emkes, and Stefan Debener. P300 speller bci with a mobile eeg system: Comparison to a traditional amplifier. *Journal of neural engineering*, 11:036008, 04 2014.

[32] Francisco Velasco-Álvarez, Álvaro Fernández-Rodríguez, Francisco-Javier Vizcaíno-Martín, Antonio Díaz-Estrella, and Ricardo Ron-Angevin. Brain–computer interface (bci) control of a virtual assistant in a smartphone to manage messaging applications. *Sensors*, 21(11), 2021.

[33] Qasem T. Obeidat, Tom A. Campbell, and Jun Kong. Spelling with a small mobile brain-computer interface in a moving wheelchair. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(11):2169–2179, 2017.

[34] Hubert Cecotti and Axel Graeser. Convolutional neural networks for p300 detection with application to brain-computer interfaces. *IEEE transactions on pattern analysis and machine intelligence*, 33:433–45, 03 2011.

[35] Markus A Wenzel, Inês Almeida, and Benjamin Blankertz. Is neural activity detected by ERP-based brain-computer interfaces task specific? *PloS one*, 11(10):e0165556, 2016.

[36] Roberta Carabalona, Ferdinando Grossi, Adam Tessadri, Paolo Castiglioni, Antonio Caracciolo, and Ilaria de Munari. Light on! real world evaluation of a P300-based brain–computer interface (BCI) for environment control in a smart home. *Ergonomics*, 55(5):552–563, 2012.

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

[38] Setare Amiri, Reza Fazel-Rezai, and Vahid Asadpour. A review of hybrid brain-computer interface systems. *Advances in Human-Computer Interaction*, 2013, 2013.

[39] Andrea De Cesarei and Maurizio Codispoti. When does size not matter? effects of stimulus size on affective modulation. *Psychophysiology*, 43(2):207–215, 2006.

[40] YoungJae Song and Francisco Sepulveda. A novel onset detection technique for brain–computer interfaces using sound-production related cognitive tasks in simulated-online system. *Journal of neural engineering*, 14(1):016019, 2017.

[41] MS Windows NT kernel description. https://github.com/vlawhern/arl-eegmodels. Accessed: 2022-03-10.

[42] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of Neural Engineering*, 15(5):056013, 2018.

[43] Schirrmeister Robin Tibor, Springenberg Jost Tobias, Fiederer Lukas Dominique Josef, Glasstetter Martin, Eggensperger Katharina, Tangermann Michael, Hutter Frank, Burgard Wolfram, and Ball Tonio. Deep learning with convolutional neural networks for eeg decoding and visualization. *Human Brain Mapping*, 38(11):5391–5420.

[44] Benjamin Blankertz, Steven Lemm, Matthias Treder, Stefan Haufe, and Klaus-Robert Müller. Single-trial analysis and classification of ERP components:a tutorial. *NeuroImage*, 56(2):814–825, 2011.

[45] Dennis J McFarland, William A Sarnacki, and Jonathan R Wolpaw. Brain–computer interface (BCI) operation: optimizing information transfer rates. *Biological psychology*, 63(3):237–251, 2003.

[46] Qi Li, Shuai Liu, Jian Li, and Ou Bai. Use of a green familiar faces paradigm improves p300-speller brain-computer interface performance. *PLOS ONE*, 10(6):1–15, 06 2015.

[47] Min-Ho Lee, John Williamson, Dong-Ok Won, Siamac Fazli, and Seong-Whan Lee. A high performance spelling system based on EEG-EOG signals with visual feedback. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(7):1443–1459, 2018.

[48] Nojun Kwak and Chong-Ho Choi. Input feature selection by mutual information based on parzen window. *IEEE transactions on pattern analysis and machine intelligence*, 24(12):1667–1671, 2002.

[49] Jerome H Friedman. Regularized discriminant analysis. *Journal of the American Statistical Association*, 84(405):165–175, 1989.

[50] Jennifer Y Bennington and John Polich. Comparison of P300 from passive and active tasks for auditory and visual stimuli. *International Journal of Psychophysiology*, 34(2):171–177, 1999.

[51] Lawrence Ashley Farwell and Emanuel Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical Neurophysiology*, 70(6):510–523, 1988.

[52] Jing Jin, Brendan Z Allison, Tobias Kaufmann, Andrea Kübler, Yu Zhang, Xingyu Wang, and Andrzej Cichocki. The changing face of P300 BCIs: a comparison of stimulus changes in a P300 bci involving faces, emotion, and movement. *PloS one*, 7(11), 2012.

[53] Zhaohua Lu, Qi Li, Ning Gao, and Jingjing Yang. The self-face paradigm improves the performance of the P300-speller system. *Frontiers in Computational Neuroscience*, 13, 2019.

[54] Teng Cao, Feng Wan, Chi Man Wong, Janir Nuno da Cruz, and Yong Hu. Objective evaluation of fatigue by EEG spectral analysis in steady-state visual evoked potential-based brain-computer interfaces. *Biomedical Engineering Online*, 13(1):28, 2014.

[55] Zhenghui Gu, Zhubing Chen, Jintao Zhang, Xichun Zhang, and Zhu Liang Yu. An online interactive paradigm for P300 brain–computer interface speller. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(2):152–161, 2019.