Investigation of EEG-based indicators of skill acquisition as novice participants practice a

lifeboat manoeuvering task in a simulator

by

© Rifat Md. Bakhtear Biswas

A Thesis submitted to the

School of Graduate Studies

in partial fulfillment of the requirements for the degree of

Master of Engineering

Faculty of Engineering and Applied Science

Memorial University of Newfoundland

Time for Convocation: October, 2018

July, 2018

St. John's, Newfoundland

ABSTRACT

Adequate training is essential in safety critical occupations. Task proficiency is typically assessed through relevant performance measures. While such measures provide information about how effectively an individual can perform the task, they give no insight about their comfort level. Ideally, individuals would be capable of executing tasks not just at a certain level of performance, but also with confidence and a high degree of cognitive efficiency.

Neural signals may provide information regarding a trainee's task proficiency that performance measures alone cannot. The purpose of this study was to investigate patterns in neural activity that are indicative of task proficiency. Ten novice participants completed ten trials of a manoeuvering task in a high-fidelity lifeboat simulator while their neural activity was recorded via 64-channel EEG. Power spectral features were used along with linear discriminant analysis to classify the data from pairs of consecutive trials. Repeated measures mixed model linear regression showed that on average, the classification accuracy of consecutive trials decreased significantly over the course of training (from 82% to 73%). Since the classification accuracies reflect how different the neural activation patterns in the brain are between the trials classified, this result indicates that with practice, the associated neural activity becomes more similar from trial to trial. We hypothesize that in the early stages of the practice session, the neural activity is quite distinct from trial to trial as the individual works to develop and refine a strategy for task execution, then as they settle on an effective strategy, their neural activity becomes more stable across trials, explaining the lower classification accuracy observed in consecutive trials later in the session. These results could be used to develop a neural indicator of task proficiency.

ACKNOWLEDGEMENTS

First of all, I would like to acknowledge my sincerest gratitude to my supervisors, Dr. Sarah Power and Dr. Brian Veitch. Without their continuous guidance and encouragement this work would never have been completed successfully. I am very thankful to get the opportunity to work with this wonderful team.

Thanks a lot to my fellow graduate students: Bret Kenny and Sinh Bui for their support and the wonderful memories throughout the research work. Thanks to all of my friends and seniors for being so supportive and helpful. And I can never thank enough my parents and family members for always being there for me.

Table of Contents

List of figures

List of tables

List of Abbreviations

BCI: Brain Computer Interface

EEG: Electroencephalography

pBCI: Passive Brain Computer Interface

VE: Virtual Environment

fNIRS: Functional Near-Infrared Spectroscopy

MEG: Magnetoencephalography

ECoG: Electrocorticography

fMRI: Functional Magnetic Resonance Imaging

PET: Positron Emission Tomography

PSPs: Post-synaptic Potentials

fMRI: Functional Magnetic Resonance Imaging

PSD: Power Spectral Density

EOG: Electrooculography

ECG: Electrocardiogram

EMG: Electromyography

IAF: Individual alpha frequency

LDA: Linear Discriminant Analysis

NASA-TLX: NASA Task Load Index

VE: Virtual Environment

1 Chapter 1 : Introduction

1.1 Problem statement

A brain-computer interface (BCI) is a system that measures central nervous system (CNS) activity and converts it into artificial output that replaces, restores, enhances, supplements, or improves natural CNS output, and thereby changes the ongoing interactions between the CNS and its external or internal environment [1]. BCIs are most often based on neuro-electric signals acquired via electroencephalography (EEG). Traditionally, BCI has been focused on "active" applications. An active BCI is a system that measures intentionally-modulated brain activity, determines which of a pre-defined set of mental states the user has generated (usually by performing some specific mental tasks), and translates this into commands for controlling an external device. The objective of active BCI research has been primarily to provide a movement-free means of communication and/or environmental control for individuals with profound motor disabilities. More recently, however, BCI for applications intended for able-bodied individuals in a variety of contexts has emerged. A passive BCI (pBCI) is a system that measures ongoing, non-intentionally modulated activity from the central (and sometimes peripheral) nervous system, extracts information about the user's cognitive and/or emotional state, and uses it to adapt human-computer interaction [2].

Passive BCIs have various potential applications for augmenting or improving existing systems. One example application is to enhance online or virtual environment-based training programs to provide users with an individualized/adaptive learning experience. Currently, training is typically assessed through behavioural performance metrics, but this gives no indication of neural efficiency of task performance. Being able to perform a task very well (i.e., with high effectiveness) but with a lot of effort is very different from being able to maintain the same level of performance with low effort (i.e., with high cognitive efficiency), particularly in hazardous, safety-critical work environments. Ideally, a good training program would ensure that the participants get up to a point where they can perform required tasks with both high effectiveness and high efficiency, so that they are more likely to be ready to execute the task in the real world where conditions may be stressful or unpredictable. Neural signals may be able to provide the information regarding cognitive efficiency, which could then be used to adapt the training program for individual users via a passive BCI. First, we need to identify neural indicator(s) of task proficiency.

Simulation based-training in virtual environments is often used in situations where "real-life" training is impractical, or even impossible, due to ethical, logistical, or financial constraints. Such training protocols could benefit from a passive BCI of this type. Previous literature has explored this for flight simulators and air traffic management [3][4][5]. However, this previous research has focused on longer term training, over weeks or months. Sometimes training, particularly some emergency response training, is done over shorter periods, even a single session, with recommended follow-up, or "refresher", sessions spaced by months or years. Also, the tasks studied thus far in the literature have

been limited, so there is a need to explore different types of tasks. It may be that a task specific neural proficiency measure may be required rather than a universal measure.

We will address these gaps in the literature by looking at a previously unstudied task (specifically, operating a lifeboat) and looking for neural indicators of skill acquisition over a relatively short, one session period.

1.2   Research Objectives

The long-term objective of this research is to develop a passive BCI for incorporation into VE-based training simulators to provide an objective, cognitive-based measure of task proficiency/learning. The short-term objective of this thesis was to identify EEG-based neural indicator(s) of task proficiency/learning over a short period of practice of a cognitive motor task performed in a VE-based training simulator.

1.3   Thesis Organization

The remainder of this thesis is organized in five chapters: literature review, methodology, results, discussion, and conclusions.

Chapter 2 presents the literature review which gives an overview of passive BCI, human learning, cognitive efficiency and their relevance to this work. It describes different methods to acquire neurophysiological data (e.g., fMRI and EEG), and their strengths and weaknesses. This section also reviews the previous applications of EEG for training or learning evaluation purposes in the literature.

Chapter 3 presents a detailed description of how the experiment was done and how the data were collected from all participants. This chapter also describes how the data were analyzed.

Chapter 4 shows the results of the data analysis, while Chapter 5 discusses these findings in more detail and identifies some limitations of the work.

Chapter 6 summarizes the main findings of this thesis and discusses some potential directions for future work.

2 Chapter 2 : Literature review

2.1 Brain computer interface (BCI)

According to Brunner et al. (2015), a brain-computer interface (BCI) is defined as "a system that measures CNS activity and converts it into artificial output that replaces, restores, enhances, supplements, or improves natural CNS output and thereby changes the ongoing interactions between the CNS and its external or internal environment" [1]. They further describe these five application scenarios as follows:

(1) BCIs can replace natural CNS output that has been lost as a result of injury or malady. Examples embody communication (through a spelling system and voice synthesis) and motorized wheelchair control.

(2) BCIs can restore lost natural CNS output. Examples include useful electrical stimulation of muscles in a paralyzed person and stimulation of peripheral nerves to revive bladder function.

(3) BCIs can enhance natural CNS output. Examples include observance of brain activity throughout prolonged demanding tasks such as driving a car and detecting lapses of attention, that alerts the person and restores attention.

(4) BCIs can supplement natural CNS output. Examples include providing a robotic arm to someone and providing a selection function for people using a joystick.

(5) BCIs can improve natural CNS output. Examples include employing a BCI in stroke rehabilitation that detects and enhances signals from a damaged cortical area to stimulate arm muscles or an orthosis to boost arm movements.

For many years, BCI research was focused on the first application listed above; that is, providing a movement-free means of communication and environmental control for individuals with severe motor disabilities. Recently, however, the field has expanded to include other medical or rehabilitation-related applications, as well as applications intended primarily for use by healthy/able-bodied users. In most cases, the research is in its very early stages and work is focused on demonstrating feasibility of application cases, establishing techniques for detecting brain states etc.

The main components of a typical BCI systems are depicted in Figure 2.1. First, the brain signals are acquired from the user via some functional imaging technique. Next the brain activity is interpreted and translated by the BCI system into an appropriate output to the application. The user observes the resulting change in the application, and this feedback modifies the experience of the user, which in turn affects their brain activity, and this continues in a positive feedback loop.
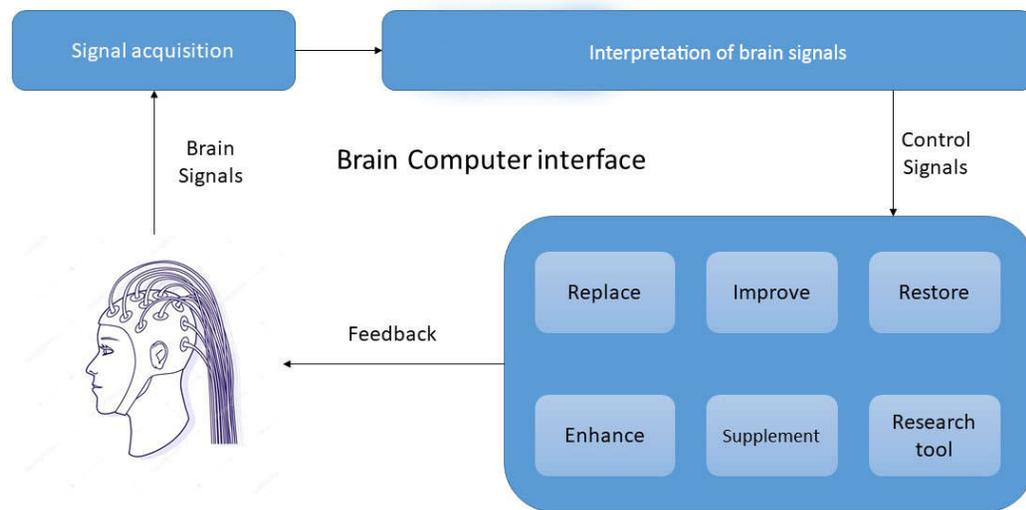
Figure 2. 1 - Principle of a brain-computer interface including possible application scenarios (adapted from [6]).

2.1.1 Active and Passive BCIs

There are a number of ways BCIs can be categorized. One way is according to the type of brain activity the BCI detects, or the way the BCI output is controlled. By these criteria, there are two main types of BCI: active and passive [1]. These can be described as follows:

a) Active BCI: An active BCI derives its outputs from brain activity which is directly and consciously controlled by the user, independently from external events, for the purpose of controlling an application. The user usually generates the desired brain activity by performing a particular mental task such as motor imagery, or mental arithmetic [7].

b) Passive BCI: A passive BCI derives its outputs from arbitrary brain activity arising without intentional or voluntary control by the user, for the purpose of enriching a

human-computer interaction with implicit information on the user's mental state (e.g., cognitive or emotional) [8].

A schematic overview of active and passive BCI is depicted in Figure 2.2.
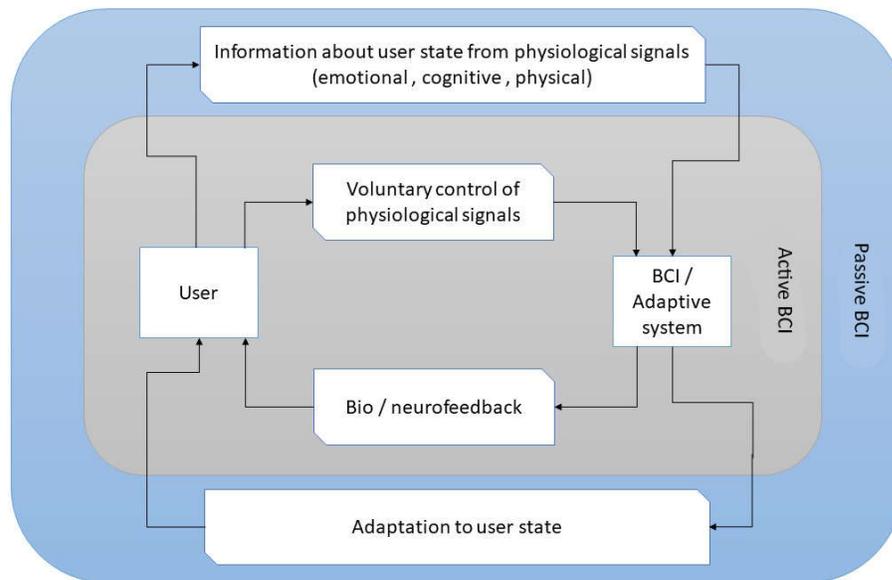
Figure 2. 2 - Schematic overview of an active and a passive BCI (Adapted from [2])

The most extensively studied application of active BCI systems is to provide a method for individuals with severe motor impairment to control external devices (e.g., computer, wheelchair) without movement of any kind. For individuals who are nonverbal and have no reliable and reproducible motor ability (e.g., late stage amyotrophic lateral sclerosis, or ALS) active BCIs could have a profound positive impact on quality of life as they may be the only hope for maintaining a means of communication with loved ones and healthcare providers. Active BCIs may also facilitate hands-free control of machines. For

example, BCI assistive robots could offer support for disabled users in daily and professional life, increasing their cooperation in building their community [5]. Furthermore, entertainment and gaming applications (e.g., [9]) have opened the market for non-medical active BCIs targeted at the general population.

Passive BCIs essentially perform user state monitoring and use the information to adapt some human-computer interface system (e.g., video game, autopilot system, graphical user interface) according to the estimated emotional or cognitive state of the user [10]. For example, a passive BCI integrated into a video game could monitor the "mental workload" of the player, and automatically make the game harder if the mental workload is too low and easier if it is too high. Much passive BCI research to date has been focused on improving safety in high-risk and safety-critical occupations like air traffic controllers, pilots, and industrial operators by monitoring states like mental workload, stress, and fatigue [11, 12]. Passive BCIs could also find application in the development of smart environments, in emotion-controlling applications, and in neuromarketing [13].

2.1.2   Invasive and non-invasive BCIs

BCIs can also be categorized according to the way the brain signals are acquired, that is, the type of functional imaging technology that is used. We can divide them into non-invasive and invasive, which are described as follows:

a) Non-invasive BCI: The sensors are placed on the scalp to measure the electrical potentials (electroencephalography, EEG), magnetic field (magnetoencephalography, MEG), or hemodynamic response (functional

magnetic resonance imaging, fMRI; functional near-infrared spectroscopy, (fNIRS) produced in the brain.

b) Invasive BCI: The sensors are placed into the cortical tissue, measuring the activity of a single neuron (intra-cortical electrodes) or populations of neurons (electrocorticography, ECoG).

While invasive BCIs work much better due to the significantly higher quality of the brain signals acquired, there are very few applications/situations in which a user would be willing to undergo implantation of sensors onto/into their brain. Therefore, for a majority of BCI applications, the focus is on non-invasive functional imaging technologies. Electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS) are the leading non-invasive BCI modalities in terms of cost and portability [14]. Recently researchers have studied combining these two modalities for improved BCI performance in what has been termed a "hybrid BCI" [15, 16].

Figure 2.3 shows a list of Invasive and non-invasive methods.
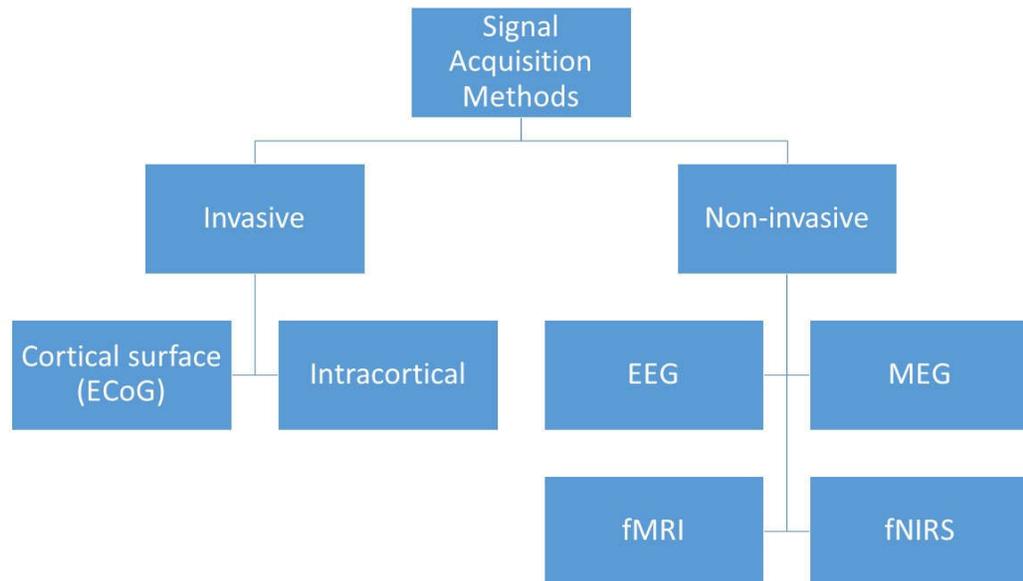
Figure 2. 3 - Signal acquisition methods

2.2 Advantages and disadvantages of EEG for BCI applications

EEG records the electric fields generated by neural activity. It has many advantages compared to alternative imaging techniques or pure behavioral observations. The most central benefit of EEG is its excellent time resolution, that is, it can take hundreds to thousands of snapshots of electrical activity across multiple sensors within a single second. This renders EEG an ideal technology to study the precise time course of cognitive and emotional processing underlying behavior. On top of that, EEG systems are relatively inexpensive, user-friendly, small, and portable, making them very practical for use in BCI applications. However, EEG does have its disadvantages as well. The use of gel and paste can be really time consuming and interference from myogenic and electric sources of noise can really disrupt the signal. For this reason, research is ongoing to eliminate these disadvantages (e.g., development of dry/wireless electrode system) in the future.

Magnetoencephalography (MEG) records the magnetic fields generated by neural activity. Like EEG, MEG has very good time resolution and is usually considered to capture deeper neural activity much better than EEG. However, MEG scanners are large, stationary, and expensive, and they limit subject movement. They also require heavy technical maintenance and training resources. As such, MEG is impractical for BCI applications.

Functional magnetic resonance imaging (fMRI) measures changes in blood flow associated with neural activity. Increased neural firing requires oxygen, which is delivered by blood, and the magnetic properties of oxygenated blood are different from those of non-oxygenated blood. This property is measured by fMRI as a distortion of the magnetic field generated by protons. fMRI has excellent spatial resolution, and for many applications is the "gold standard" of functional imaging. However, it lacks the time resolution of EEG and is large, expensive, and stationary, so for BCI applications it is not at all feasible.

Positron emission tomography (PET) is an invasive nuclear imaging technique based on gamma radiation of a decay that is inserted into the body of the respondent. With PET, researchers can monitor metabolic activity (for example, glucose metabolism) of neurons during cognitive activity. While PET scans are much more robust towards motion artifacts, they lack the high time resolution of EEG recordings.

2.3   Fundamentals of EEG

2.3.1   Origin of EEG

The EEG records electrical activity originating in the cerebral cortex via electrodes placed on the surface of the scalp. The electrical signals originate from neurons, cells in the brain that communicate with each other through electrical impulses [17]. There are a number of possible sources of the EEG signals, including action potentials, post-synaptic potentials (PSPs), and chronic neuronal depolarization. Action potentials induce a brief (10 ms or less) local current in the axon (the long threadlike part of a nerve cell along which impulses are conducted from the cell body to other cells) with a very limited potential field. This makes them unlikely candidates. PSPs are considerably longer (50–200 ms), have a much greater field, and thus are more likely to be the primary generators of the EEG [18].

2.3.2   Electrode placement

To aid in the interpretation of EEG signals across laboratories, studies, etc., electrode placement has been standardized according to what is known as the 10-20 International System of Electrode Placement (Figure 2.4a). Invented by Dr. Herbert Jasper at the Montreal Neurological Institute in 1958 [19], the name comes from the fact that the distances between adjacent electrodes are either 10% or 20% of the total front–to-back or right–to-left distance of the skull. This system was developed to ensure standardized reproducibility, so that data can be compared within an individual over time as well as between individuals. The names of the electrode locations in the 10-20 system are shown in Figure 2.4a and make reference to the underlying brain region. For example, electrode Fz lies over the centre of the frontal cortex, while electrode P3 lies over the left parietal

cortex. To allow the placement of a greater number of electrodes, the 10-10 system (see Figure 2.4b) can be used [18].
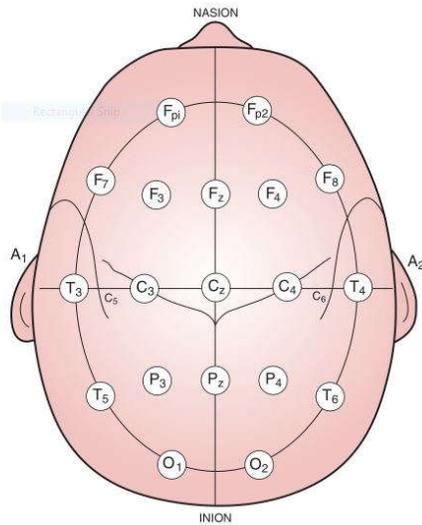
Figure 2.4 a - 10-20 system   Figure 2.4 b - 10-10 system
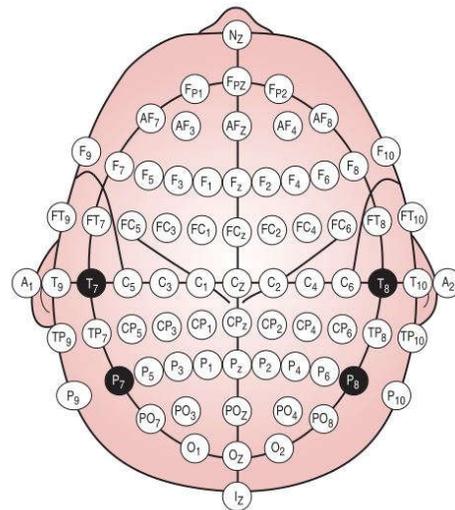
2.3.3 EEG frequency bands

The electrical impulses in an EEG recording are measured in microvolts (μV) and look like wavy lines with peaks and valleys. These EEG brainwaves change according to what we are doing and feeling. Researchers have classified these brain waves according to specific frequency bands described below. EEG signal components at particular frequencies are related to the synchronized oscillation of populations of neurons.

Delta band (varies from 1-4 Hz): Delta brainwaves are slow and deeply penetrating. They are generated in deepest meditation and dreamless sleep. Delta waves suspend external awareness and are the source of empathy. Healing and regeneration are stimulated in this state [20, 21].

Theta band (varies from 4-8Hz) : Theta brainwaves occur most often in sleep but are also dominant in deep meditation. Theta is our entryway to learning, memory, and intuition. In theta, our senses are withdrawn from the external world and focused on signals originating from within [22].

Alpha band (varies from 8-12 Hz) : Alpha brainwaves are dominant throughout quietly flowing thoughts, and in some pondering states. Alpha is the resting state for the brain. Alpha waves aid overall mental coordination, calmness, alertness, mind/body integration and learning [20].

Beta band (varies from 13-30 Hz) : Beta brainwaves dominate our traditional waking state of consciousness when attention is directed towards cognitive tasks and the outside world. Beta is a 'fast' activity, present when we are alert, attentive, engaged in problem solving, judgment, decision making, or focused mental activity [20].

Gamma band (frequency above 30 Hz) : Gamma brainwaves are the fastest of brain waves and relate to simultaneous processing of information from different brain areas. Gamma brainwaves pass information rapidly. A prominent gamma rhythm provides a signature of engaged networks. Gamma has been observed in a number of cortical areas, as well as subcortical structures, in numerous species. In sensory cortex, gamma power increases with sensory drive and with a broad range of cognitive phenomena, including perceptual grouping and attention [7].

2.4 Potential passive BCI application: Training assessment

Another potential passive BCI application is to enhance online or simulation-based training by providing a cognitive measure for assessment of skill acquisition. Inappropriate training assessment might have either high social costs or economic impacts, especially for pilots, air-traffic controllers, surgeons, and other high-risk occupations or work environments. One of the current limitations of standard training assessment procedures is the lack of information about the amount of cognitive resources required by the individual for the correct execution of the proposed task. In fact, even if the task is accomplished achieving the maximum performance, by the standard training assessment methods, it would not be possible to gather and evaluate information about cognitive resources available for dealing with unexpected events or emergency conditions. Therefore, a metric based on the brain activity (neurometric) able to provide this kind of information should be very important.

The difference between the available cognitive resources and the amount of those involved for the task execution is called *Cognitive Spare Capacity* [24]. The higher the cognitive spare capacity during a normal working activity (i.e., the operator is involving a low amount of cognitive resources), the greater the operator cognitive processing efficiency is to perform secondary tasks or to react to unexpected emergency events.

2.4.1 Cognitive Processing efficiency

Before we get to know about processing efficiency theory, we need to know cognitive interference theory. Cognitive interference refers to the unwanted and often disturbing thoughts that intrude on a person's life [24]. The central assumptions of cognitive

interference theory are that the experience of anxiety involves having various task-irrelevant thoughts (e.g., self-preoccupation; worry), and that these task-irrelevant thoughts affect performance by reducing the amount of attention available to be allocated to a central ongoing task. In this context, anxiety does not refer to the generally understood state of emotional stress associated with the term, but rather refers to the state of worry experienced by individuals when unsure how to complete a task and are aware of the associated impact on their performance.

The first problem with that theory was addressed by drawing a distinction between performance effectiveness and processing efficiency. Performance effectiveness is easily defined, because it refers to the quality of performance (e.g., as assessed by outcome measures such as accuracy and speed of task performance). Processing efficiency is based on the relationship between performance effectiveness and the amount of effort or resources used to attain that level of performance. Task-irrelevant thoughts such as worry and self-preoccupation are assumed to impair processing efficiency [25]. But as people become more comfortable with a task and develop an effective strategy for its execution, information that is relevant to a task tends to be selectively retrieved and task-irrelevant thoughts subside, resulting in increased processing efficiency [26].

2.5   Neuroimaging studies of training/learning

Research with both animal models and humans has shown that changes in neural representations can be induced not solely in response to lesions of input or output pathways, but that the organization of the adult cerebral cortex can change substantially as a result of practice and experience. Discoveries of these kinds challenge us to

investigate how it is that the brain changes in response to expertise. Modern neuroimaging methods such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) are excellent tools in this endeavour, enabling the examination of how the brain changes in response to practice or repeated exposure to a particular task.

A growing number of human functional neuroimaging studies are investigating the changes in brain activation that occur as a result of practice on a variety of motor, visuomotor, perceptual and cognitive tasks. Across studies, three main patterns of practice-related activation change can be distinguished. Practice may result in an increase or a decrease in activation in the brain areas involved in task performance, or it may produce a functional reorganization of brain activity, which is a combined pattern of activation increases and decreases across a number of brain areas [27].

Shadmehr and Holcomb (2007) used PET to examine brain activations during a task requiring individuals to control a robotic arm and perform rapid reaching movements. Practice resulted in task performance becoming highly skilled. A redistribution of activations from the frontal cortex to the posterior areas (parietal cortex, and cerebellum) was only observed during a recall session that took place 5.5 hr after the last practice session. Interestingly, these changes occurred in the absence of any further performance changes. The authors propose that this is consistent with the notion that acquisition of skilled movement is mediated through prefrontal cortex structures, and with time and practice, and as automation occurs, motor structures such as the cerebellum assume a greater role and possibly become the site of motor memory [28].

Petersson et al. (1999) also observed a reorganization of functional activations. Using PET, they compared practised to novel free recall of abstract designs. While they observed decreases in activation associated with decreased reliance on attentional and working memory processes (in areas like the prefrontal cortex, anterior cingulate cortex, and posterior parietal cortex), and increases in areas related to task-irrelevant processing (e.g., auditory cortex, insula), they also saw reorganization of activations relating to the transition from reliance on non-consolidated representations (inferior temporal cortex) to the recall of more developed representations of the abstract designs (occipito-temporal region). In other words, a reorganization of activations was observed as the task processes shifted from effortful, working memory-based recall to those involving more automatic recall of consolidated representations of the abstract designs [29].

Kassubek et al. (2001) scanned subjects using fMRI as they read either mirror inverted words or plain text before and after a training session. Reading of mirror-inverted items activated the dorsal visual pathway and premotor cortex, and a significant practice-related reduction in activation in these areas was observed [30].

Learning may also change the degree to which subjects are aware of the task or stimulus structure, which may result in a change in neural activity [27]. The changes these variables can produce can be mistakenly identified as reflecting practice-related processes rather than consequences of those processes [31].

2.5.1 EEG studies of training/learning

Gutierrez and Ramírez-Moreno (2015) used EEG to quantify changes in brain activity associated with the progression of the learning experience through the functional analysis-of-variances (FANOVA) estimators (main effect function, treatment function and zero mean gaussian errors) of the EEG power spectral density (PSD). Such functional estimators provide a sense of the effect of training in the EEG dynamics. For that purpose, they implemented an experiment to monitor the process of learning to type using the Colemak keyboard layout during a twelve-lessons training with a repetition of five times for each lesson. Their aim was to identify statistically significant changes in PSD of various EEG rhythms at different stages and difficulty levels of the learning process. A series of statistical tests were performed in order to determine the personalized frequencies and sensors at which changes in PSD occur, then the FANOVA estimates were computed and analyzed for each subject, at each difficulty level, and for the selected frequency bands. Their experimental results showed a significant decrease ($p < 0.05$) in the power of beta and gamma rhythms for ten volunteers during the learning process, and such decrease happened regardless of the difficulty of the lesson. The speed of such decay in power was approximated by the slope of the linear regression of the data. They also performed one-way ANOVA and multiple-comparison tests in order to assess the effect of repeating the lesson [32].

The aim of a study that was done by Borghini et al. (2013) was to analyze the variation of the EEG power spectra in the theta band when a novice starts to learn a new task, in particular a flight simulation task. The goal was to find out the differences from the

beginning of the training to the session in which the performance level is good enough for considering him/her able to complete the task without any problems. A total of five 30-minute training sessions were completed over five consecutive days, at the same time each day. While the novices were engaged in the flight simulation tasks they recorded the brain activity by using high resolution EEG techniques as well as neurophysiologic variables such as heart rate and eye blinks rate. They found that EEG power in the theta band at the frontal site showed an inverted U-shaped relation during the training sessions for flight simulation tasks. One possible explanation for the phenomena is that the brain first worked hard to learn how to use specific task-relevant areas, followed by improvement of efficiency derived from disuse of irrelevant brain areas for good task performance [4].

Taya et al (2015) employed the functional connectome approach to study the changes in global and local information transfer efficiency of the functional connectivity induced by training of a piloting task. In this study, they investigated training-induced topological changes of the brain functional network using the graph theoretical approach. Four levels of difficulty were defined to induce a variety of mental workload levels in the participants; hyper-easy, easy, medium, and hard. Each subject underwent five consecutive days of training, and each session consisted of two sets of the four difficulty conditions. Their results have demonstrated that global information transfer efficiency of the network, revealed by normalized characteristic path length in beta band, once decreased and then increased during the training sessions [33].

Borghini et al. (2014) analyzed the possibility of applying a neuro-electrical cognitive metrics for the evaluation of the training level of subjects during the learning of a task

employed by Air Traffic Controllers. In particular, EEG as well as ECG (electrocardiogram) and EOG (electrooculogram) signals were recorded from a group of students during the execution of an Air Traffic Management task, proposed at three different levels of difficulty [3]. From this research, they discovered that, by focusing the analysis on the direct and inverse correlation of the frontal theta (4 –7 Hz) power and heart rate, and of the parietal alpha (10-12 Hz) power and eye blink rate, respectively, with the degree of mental and emotive engagement, it is possible to obtain useful information about the training improvement across the training sessions. The subjects completed one-hour training sessions daily for five days.

Olga et al. (2016) proposed an EEG-based mental state monitoring system that can reflect the true "inner" feelings, stress level, and workload of maritime cadets during a simulator-aided assessment. They analyzed the recognized brain states and the corresponding performance and behavior recorded by the simulator to study how human factors affect the subjects' performance. For example, they managed to check if there is any correlation of the cadet's stress level and performance results. Finally, they proposed an EEG-based system that allowed them to assess whether a cadet is ready to perform tasks on the bridge or needs more training in the simulator even if he/she navigated with few errors during the assessment. The 14 channel Emotiv device [34] was used to capture the users' EEG signals. For this experiment, participants underwent four exercises on a single day where the length of each exercise was thirty minutes [35].

Liu et al. (2016) figured that EEG signals can be used to directly assess the "inner" mental state of the trainee. In this paper, they proposed an EEG-based system which can monitor

emotion, workload, emotional stress, and environmental stress for human factors evaluation in the simulator-added assessment. During their study, human factors measurements including mental workload, stress, and emotion of cadets while performing the navigation tasks were obtained in real time using the EEG device. They hope to use the proposed system to monitor and understand the brain states of the subjects during the assessment. The analysis of data and developing the system which can assess whether a cadet is well trained to perform the tasks on the bridge or needs more training is the next step of their work. For this experiment participants underwent four exercises on the same day where the exercises were of variable duration (between 12 to 23 minutes) [36].

Perry et al. (2015) used a dental haptic simulator to figure out the cognitive difference between a group of novice and expert surgeons and figured that experts have a lower T3-Fz coherence than novices. Seven 'experts' (qualified dentists) and eight 'novices' (first year dental students) volunteered for the study. Each subject was given a set time to complete a clinical and a non-clinical task on a dental haptic simulator. EEG was recorded in the T3, T4 and Fz regions for coherence assessment. Although there was no significant difference in the performances of experts and novices, a trend was evident toward lower T3-Fz and possibly T4-Fz coherence in experts when compared to novices [37].

Borghini et al. (2015) worked to figure out if it was possible to obtain quantitative information about the degree of the learning process throughout a training period by analyzing the variations in EEG, ECG, and EOG. A group of ten subjects trained daily with the NASA multi-attribute-task-battery (MATB). During such training period, physiological, behavioral, and subjective data were collected and analyzed. Their results

suggested that the EEG signals changed consistently across the training sessions and that they have correlations with the overt behavior of the subjects. They observed a clear increase of the frontal theta power from the first day of the training to the third day, and a decrease from the third to the fifth day. The same trend, but with opposite sign, was observed for parietal alpha power [38].

Borghini et al. (2017) collected EEG data and the performance of ten participants along a training period of three weeks, while learning to execute a NASA-MATB task. Specific indexes were estimated from the behavioral data and EEG signal to objectively assess the users' training progress. They proposed a neurometric based on a machine learning algorithm to quantify the user's training level within each session by considering the level of task execution, and both the behavioral and cognitive stabilities between consecutive sessions. The proposed neurometric takes into account the mean performance level achieved by the user (capability in executing the task correctly), the stability of the performance across different sessions (capability in maintaining high performance over time), and the stability of the brain activations across consecutive training sessions (capability in dealing with the task requiring the same amount of cognitive resources once it became automatic). By considering such aspects, they proposed a measure of the training level and to assess if the single user could be considered "trained" or not [39].

All of these studies looked at some certain trend over time, and their experiments consisted of training of various tasks over multiple sessions over multiple days. In this thesis, the objective was to explore whether observable trends exist in the EEG signals over a much shorter training period, specifically a single session. Moreover, in the aforementioned

studies, analysis was done on single brain region/frequency band combinations individually. In this work, an approach was taken that would allow the consideration of multiple frequency bands and regions simultaneously. Finally, this work involves a previously unstudied task.

2.6 Virtual environment for training

Virtual training environments are used when real-life training is challenging because of the high costs, danger, time, or effort involved. Training is a promising application area of three-dimensional virtual environments. These environments allow the trainees to navigate through and interact with a virtual representation of a real environment in which they have to learn to carry out a certain task [40].

Training in virtual environments can be very effective. Rose et al. (2000) conducted a study to measure and evaluate skill transfer of a simple sensorimotor task trained in a VE to real world performance. They found that virtual and real training resulted in equivalent levels of post-training performance, both of which significantly exceeded task performance without training. They also found that real task performance after training in a VE was less affected by concurrently performed interference tasks than was real task performance after training on the real task [41].

3   Chapter 3: Methods

3.1   Experimental overview:

A total of 14 volunteers participated in this experiment. Participants were asked to perform a task in a high-fidelity lifeboat simulator developed by Virtual Marine (St. John's, Canada) while their EEG signals were recorded. This chapter will first describe the design of this experiment and then focus on the assessment of EEG-based skill acquisition indicators that will help us to understand the change in brain signal as a participant improves in a single task.

3.2   Study participants

A total of 14 healthy participants (all male, mean age: 29 ± 4.85 (mean $\pm$ std)) were recruited on a volunteer basis from the general population at Memorial University of Newfoundland. Only male participants were included to reduce the effect of any sex-related differences in brain function or size [42, 43, 44].

Participants were included in the study if they met the following criteria:

1) Were 18-65 years of age,

2) Had normal vision, or vision corrected-to-normal via contact lenses (eyeglasses can impede adequate EEG electrode contact with the scalp),

3) Had normal or correct-to-normal hearing,

4) Had no history of neurological disorder, disease, or injury and no cognitive impairment,

5) Had no prior experience operating a watercraft driven with a steering wheel and throttle, either in a real or virtual-environment.

As a part of the preparation for the experiment, participants were asked to refrain from exercising, smoking, or consuming caffeine or alcohol for at least four hours prior to the session. To maximize signal quality, they were also asked to wash their hair on the day of the experiment, and refrain from using hair products other than regular shampoo.

Information regarding age, sex, handedness, driving experience, and gaming experience of the participants were collected at the beginning of the session (it was thought that individuals with more gaming and/or driving experience may perform better at the lifeboat task than others).

All participants gave written informed consent to participate in this study. The experimental design was approved by the Interdisciplinary Committee on Ethics in Human Research at Memorial University of Newfoundland.

3.3   Instrumentation

Figure 3. 1 - ActiChamp Amplifier with ActiCap active electrodes

Neural data were collected via a 64-channel EEG system (ActiCHamp amplifer with ActiCAP active electrodes, Brain Products GmbH; see Figure 3.1) at a sampling rate of 500 Hz. Electrodes were placed according to the international 10-20 system and secured via a flexible cap (see Figure 3.2).
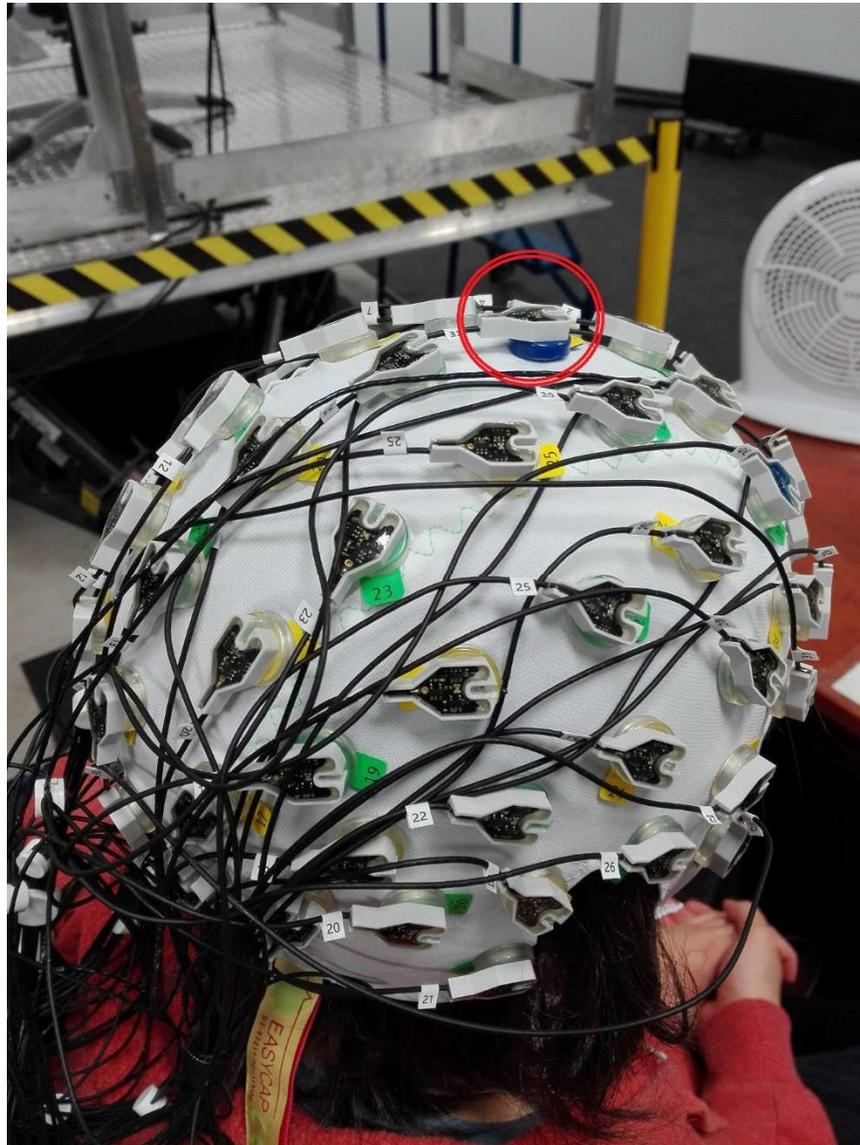
Figure 3. 2 - Participant wearing the electrode cap. Electrode FCz is circled.

Electrolyte gel was used to achieve good coupling of each electrode to the scalp, with the objective being to reduce the electrode impedance to ≤ 10 kΩ. Electrode FCz (circled in Figure 3.2) was used as the reference, and the ground electrode was placed in the centre of the forehead.

In addition to the EEG data, electrooculography (EOG) and electrocardiogram (ECG) data were collected via various sensors placed on the hands, face, and chest and connected to the auxiliary input channels of the ActiCHamp. Analysis of these data is not included in this study. Brain Vision Recorder software was used to record and save the EEG and other physiological signal data.

3.4 Experimental design

3.4.1 Lifeboat simulator and task

In this study, we used a lifeboat simulator developed by Virtual Marine Technology (St. John's, Canada). The simulator was developed to train lifeboat coxswains in proper lifeboat evacuation procedures. The simulator provides a very realistic first-person view as if the participant is inside the lifeboat itself. The simulator does not provide any motion simulation, but it is built to include original operating controls from lifeboat manufacturers, realistic virtual models of specific production facilities or drilling rigs, evacuations in emergency situations such as fire, collisions, and loss of stability, and weather and sea condition. Two large computer screens serve as the front "windshield", showing a realistic front view, while smaller screens on either side display what would be seen out of the "side windows" of the lifeboat. The systems are customized, so coxswains learn to use lifeboats while launching from offshore facilities, based on emergency evacuation plans. Individuals can be trained in lifeboat launching procedures (e.g., from a larger vessel or oil platform), as well as in driving the lifeboat in the water. The high-fidelity simulator is capable of simulating a wide range of weather and sea conditions, as well as a variety of training scenarios with varying complexity.
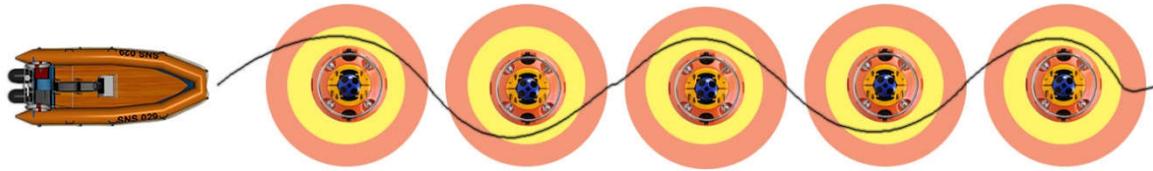
Figure 3. 3 - Schematic representation of the lifeboat task

For the purposes of this study, we selected a simple manoeuvering task with which a novice participant could reach a reasonable level of competence within a single 2-3 hour experimental session (including equipment setup). Specifically, participants were asked to manoeuver the lifeboat, using a steering wheel and throttle, in a "zig-zag" motion from a starting point through a series of five buoys (i.e., a slalom course, see Figure 3.3). The buoys were spaced 36 meters apart, and wave conditions were set to a "ripple". Weather conditions were sun and clouds with a light 2 km/h wind velocity. These task parameters were determined via pilot testing to result in a task of appropriate difficulty (i.e., pilot participants generally started out with very poor performance but reached a high level of performance based on objective behavioural measures within a 1-2 hour practice period).

3.4.2 Experimental protocol

Participants completed a single experimental session that lasted approximately 2-3 hours, including equipment setup. The actual experimental protocol generally took between 1-2 hours, including breaks. After the experimenter described the study and obtained informed

consent, the participants were positioned inside the lifeboat simulator and equipped with the EEG and other sensors. They were given a few moments to become comfortable with the simulator and the controls (throttle and steering wheel). Participants were asked to minimize movement, particularly of the head, as much as possible during the trials in order to reduce signal artifacts.

Participants completed a total of ten trials of the lifeboat manoeuvering task. Participants were instructed to maintain the zig-zag path as accurately as possible, keeping the lifeboat close to the buoys without hitting them. They were also told not to go back if they missed a buoy, but rather to keep going through the course. There were no time constraints, and participants generally took 3-5 minutes to complete each trial. Participants were given general feedback after each trial about how they did (e.g., you hit one buoy at high speed), but were not given any advice on how to improve their performance; it was left to them to figure out a strategy that worked best for them.

After each trial, participants were asked to complete an abbreviated version of the NASA-TLX questionnaire [45] to provide a measure of the perceived workload they experienced during the trial. This took approximately 1-2 minutes each time. The NASA-TLX is a common method of assessing individuals' perception of the workload experience while performing a task. It incorporates multiple dimensions of workload including mental, physical, and temporal demand, performance, effort, and frustration. We expected that the participants' perceived workload would decrease as they became more comfortable with the task. In addition, the participants were asked to provide a separate rating of the

difficulty of the task during the preceding trial, on a scale from 1 (very easy) to 100 (very difficult).

Prior to each lifeboat trial, two baseline trials of one minute each were recorded, one with eyes open and one with eyes closed. Data from these trials were used to calculate the Individualized Alpha Frequency (see section 3.8.1) as well as to perform baseline normalization of the EEG data (see Section 3.7). A combination of one eyes closed baseline, one eyes open baseline, and one lifeboat trial comprised one experimental "block", of which there were ten total. Participants were allowed to rest as needed between blocks.

Figure 3.4 shows the timeline of the experimental session.
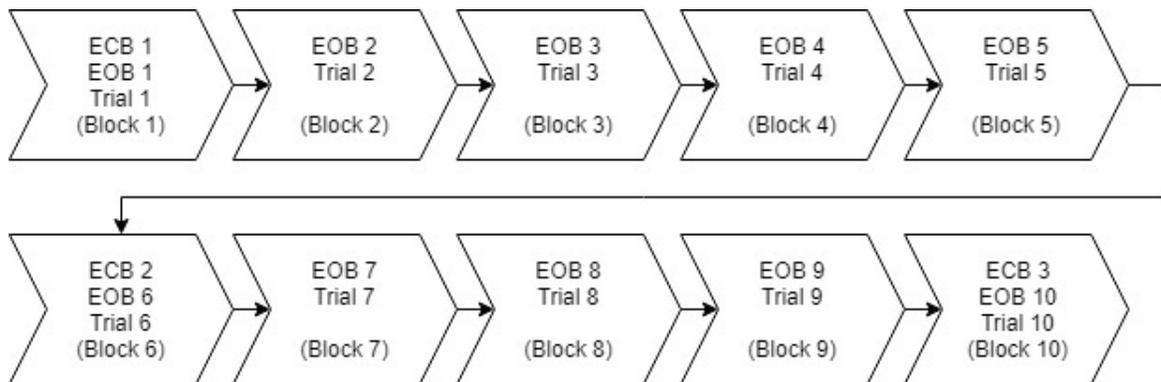
Figure 3. 4 - Timeline of experimental session (EOB = Eye Open Baseline, ECB = Eye Close Baseline)

3.5   Calculation of performance score

A performance metric was derived to provide an objective measure of each participant's performance during each trial. This measure was derived based on advice from experts at Virtual Marine, and incorporates the speed, accuracy, and control with which the

participants' manoeuvered through the lifeboat course. Speed was indicated simply by the time taken to complete the trial. Accuracy and control were indicated by how tightly the participant could manoeuver around the buoys without hitting them. Specifically, the following information was used to derive the performance score:

1) Number of buoys hit (i.e., were "bumped into" by the lifeboat)

2) Speed of contact during a buoy hit (a "hard" hit was considered to be at ≥ 9 knots while a "soft" hit was at < 9 knots)

3) Number of buoys skipped (i.e., were not fully manoeuvered around)

4) Total distance travelled from start point to finish point

5) Total time taken (in seconds) to complete the trial

The performance score was calculated as follows:

$$\text{Performance} = \frac{accuracy\ score - (6\ x\ hard\ hit) - (3\ x\ soft\ hit)}{distance\ x\ time\ in\ seconds}\ x\ 1000 \qquad \text{(Equation 3.1)}$$

In this equation, for each "hard" buoy hit, six points were deducted and for each "soft" buoy hit three points were deducted. Performance score was inversely proportional to both distance and time. Participants could get at most 25 points for their "accuracy score": for each of the five buoys, they would get no points if they did not travel within 18 meters of the buoy, three points if they travelled between 9 and 18 meters from the buoy, and five points if they stayed within 9 meters of the buoy. The circles in the schematic diagram of Figure 3.3 represent these zones.

3.6   EEG data analysis.

3.6.1   Data pre-processing

One challenge of EEG data is its high signal-to-noise ratio. The EEG signal can be contaminated with other electrophysiological artifacts including EOG (from eye movements), ECG (from the heart), and EMG (electromyography, from muscle activity), as well as electrical noise from the environment, or simply movement of the electrodes. As such, the EEG data must undergo pre-processing to remove any unwanted artifacts. EEGLAB is a MATLAB toolbox that was used to do the initial processing on the EEG data. EEGLAB is distributed under the free GNU GPL license for processing data from EEG other electrophysiological signals. Along with all the basic processing tools, EEGLAB is able to implement independent component analysis (ICA), time/frequency analysis, artifact rejection, and several modes of data visualization. This MATLAB toolbox for EEG data analysis was downloaded from this website: https://sccn.ucsd.edu/eeglab/index.php.

The raw EEG data was imported into EEGLAB from Brain Vision Recorder for pre-processing. Figure 3.5 shows a sample of the raw EEG from the 64 EEG channels, as well as the other collected physiological signals (e.g., ECG at bottom).
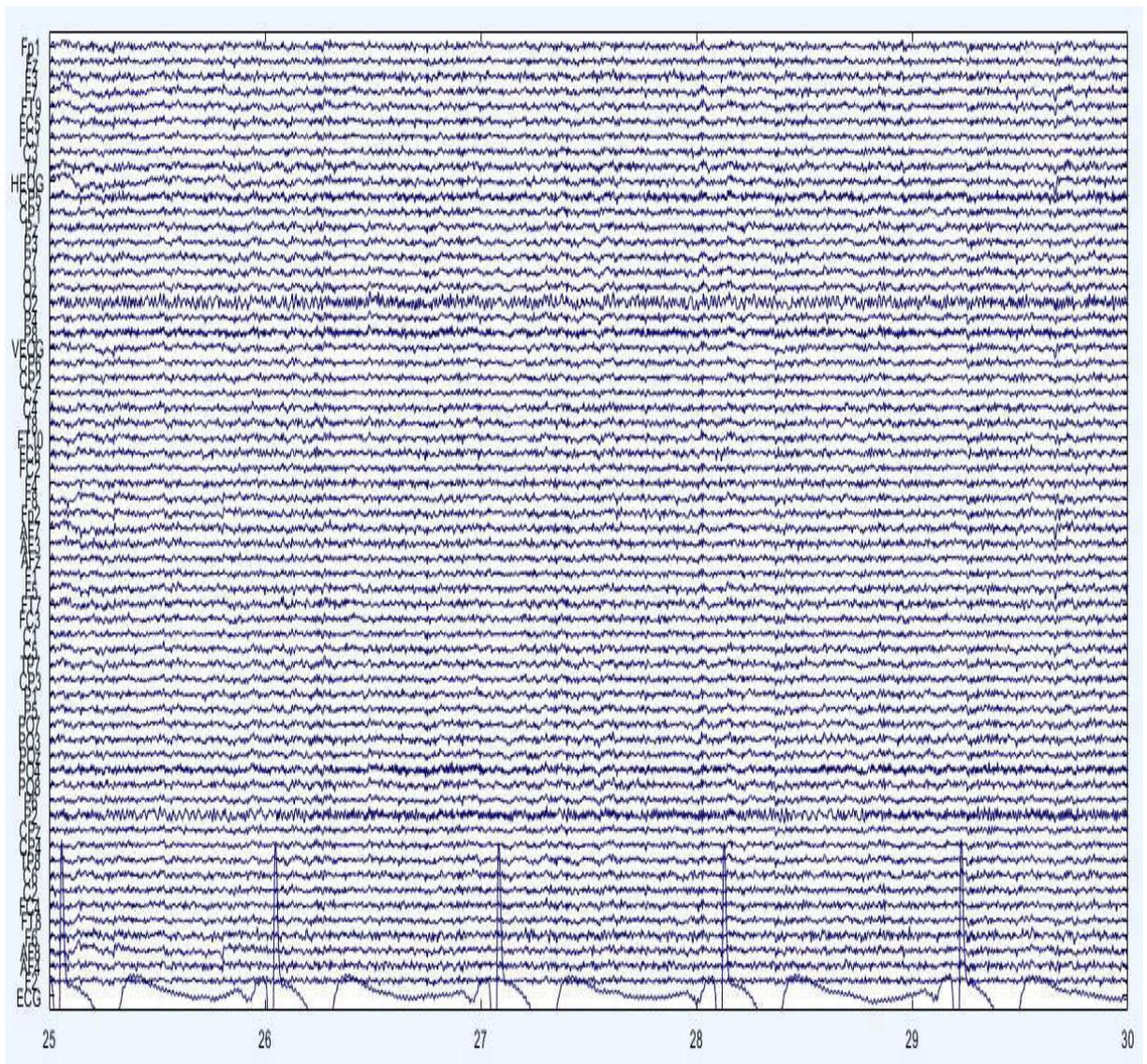
Figure 3. 5 – Sample raw EEG data from 64 electrodes. Includes EOG and ECG also.

3.6.2 Removal of EMG artifact

The first step in the noise removal was to manually remove the EMG noise due to muscle activity (e.g., from clenching of jaw, forehead, etc.). EMG artifacts are hard to remove using any algorithm. So, in order to minimize such artifacts, participants were instructed to try to refrain from moving as much as possible during both lifeboat and baseline trials.

However, given the nature of the task, movement could not be completely avoided. Consequently, we manually removed the small segments of data that were contaminated with motions artifact. Figure 3.6 shows a sample of EEG data contaminated by EMG artifact.
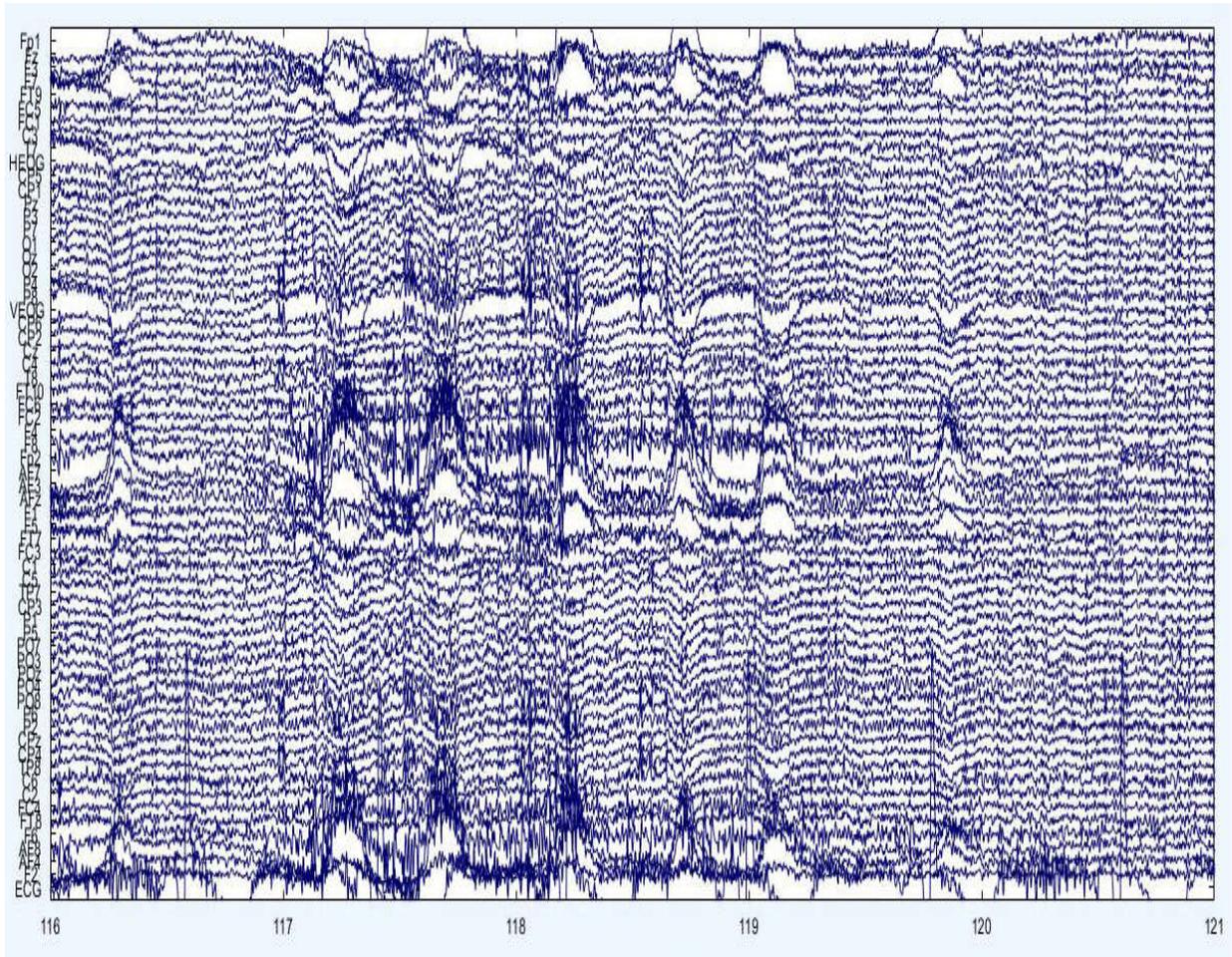
Figure 3. 6 - EEG signals contaminated by EMG noise

3.6.3 Removal of low and high frequency content

Most of the salient information within the EEG signal is contained in the frequency band 1 – 40 Hz, therefore we applied a Chebyshev Type 2 bandpass filter to retain these

frequencies. A Chebyshev Type 2 filter was selected due to its relatively steep roll off in the transition band as well its maximally flat passband response.

3.6.4   Removal of eye movement artifact

Some of the most prominent artifacts present in the EEG signal are due to eye blinking and side-to-side and up-and-down eye movements (saccades). Independent component analysis (ICA) is perhaps the most popular method of eye artifact removal in EEG signal processing. Independent component analysis (ICA) is a computational method for separating a multivariate signal into additive subcomponents. ICA is a special case of blind source separation [46]. ICA attempts to decompose a multivariate signal into independent non-Gaussian signals. It is very effective for isolating subcomponents associated with the eye blinks and the saccades, which can be identified visually and removed before reconstructing the signal. Figure 3.7 shows an example of eye artifact removal via ICA. Figure 3.7a depicts the original signal contaminated with eye blink and saccades. Figure 3.7b shows the signal decomposed into independent components; components 8, 17 and 21 contain the eye blinks and saccades, respectively. Finally, Figure 3.7c shows the signal reconstructed after removal of those components.
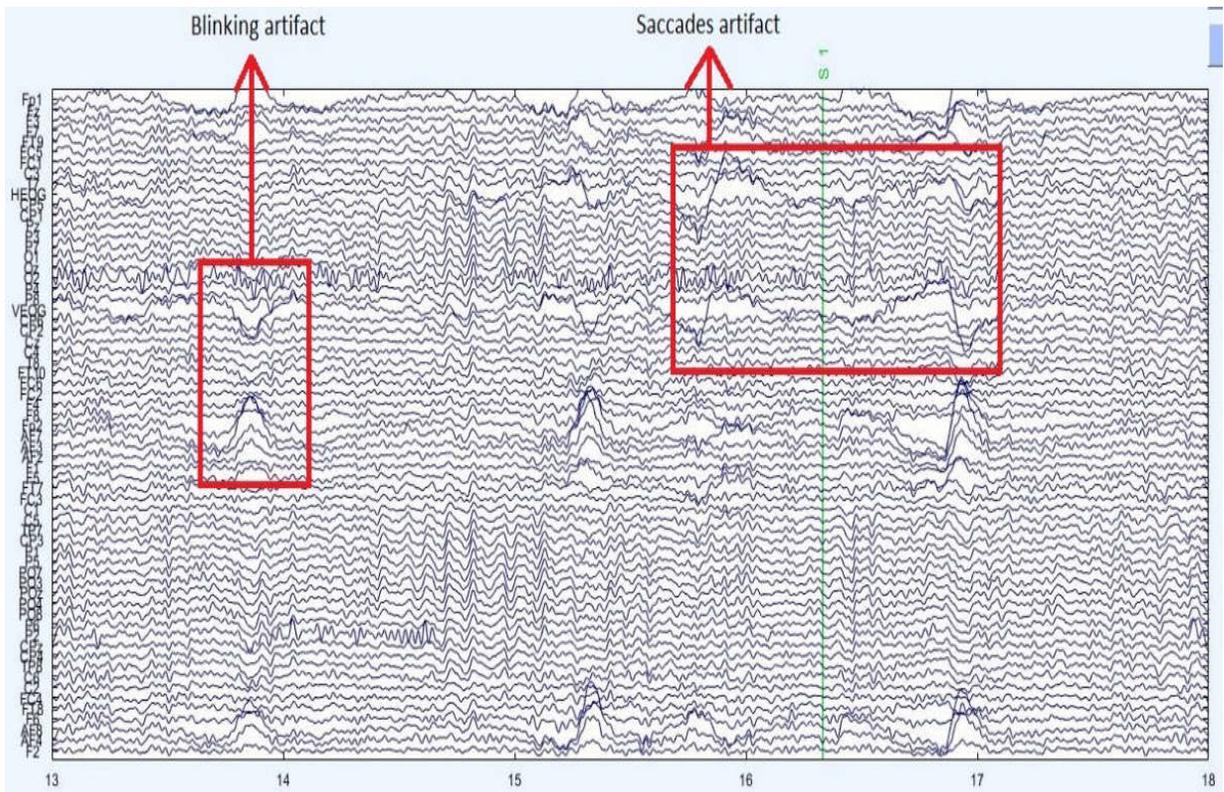
Figure 3. 7a - EEG data contaminated with blink and saccades artifact
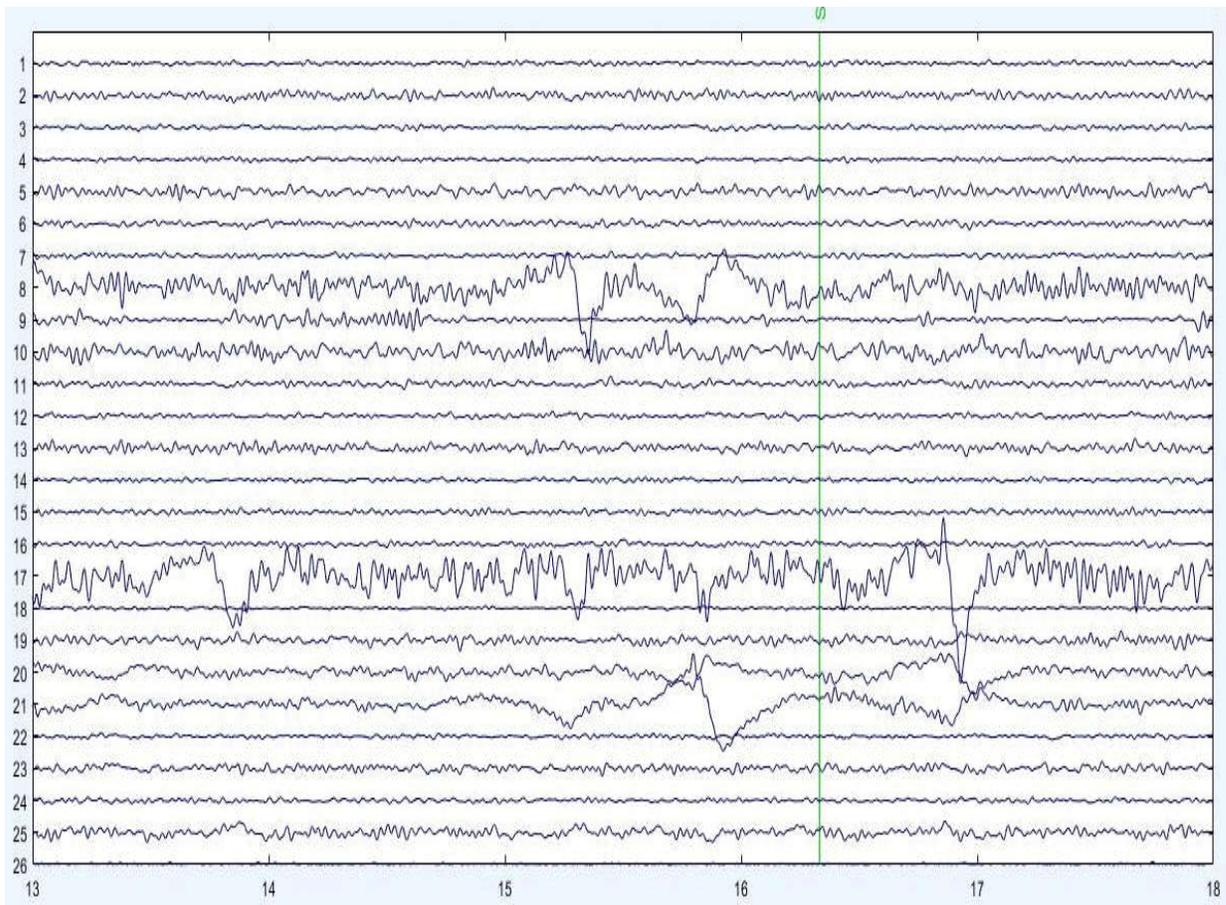
Figure 3.7 b - Independent components from ICA. Components 8, 17 and 21 represent the blinks and saccades.
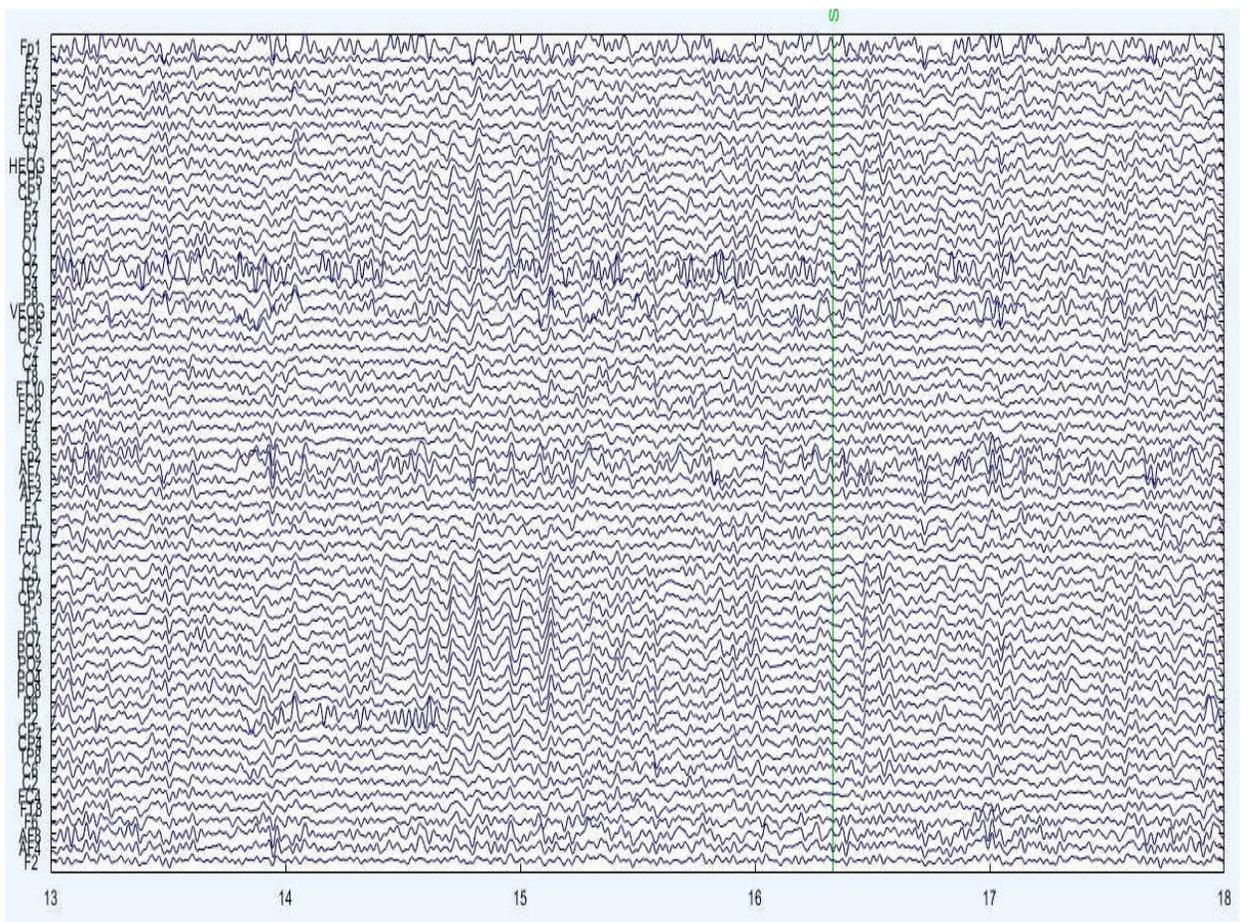
Figure 3.7 c - EEG data with blink and saccade components removed.

3.7 Baseline normalization

Baseline normalization was done on the EEG data from the lifeboat trials. The reason for doing this is that the frequency spectrum of data tends to show decreasing power at increasing frequencies. This is not specific to EEG data, but also characterizes the relationship between power and frequency of many signals, including radio, radiation from the Big Bang, natural images, and many more [47]. This decrease in power as a function of an increase in frequency follows a "1/ f" shape. This is why it is difficult to visualize activity from a large range of frequency bands simultaneously. The 1/f

phenomenon entails five important limitations to interpreting and working with time-frequency power data:

1) It is difficult to visualize power across a large range of frequency bands,

2) It is difficult to make quantitative comparisons of power across frequency bands,

3) Aggregating effects across subjects can be difficult with raw power values,

4) Task related changes in power can be difficult to disentangle from background activity,

5) Raw power values are not normally distributed because they cannot be negative and they are strongly positively skewed.

In addition, the experimental sessions were generally around 2-3 hours in duration, including time for instructions and equipment setup, so it is possible that later baseline and trials data were affected by factors such as fatigue. To eliminate such factors, the signal from each lifeboat trial was normalized by the eyes-open baseline trial immediately preceding it via the z-transform as follows:

$$Z_{tf} = \frac{trial_{tf} - \overline{baseline_f}}{n^{-1} \sum_{i=1}^{n}(baseline_{if} - \overline{baseline_f})^2} \qquad \text{(Equation 3.2)}$$

where n is the number of time points in the baseline period, the horizontal bar over baseline indicates the mean across the baseline time period, and t and f are time and frequency points. The denominator in this equation is the formula for the standard deviation of the baseline period. This normalization was done individually for each frequency band/ electrode combination.

3.8 Assessing stability of brain activity via statistical classification of consecutive trials

The main objective of this thesis was to identify a neural indicator of learning/task proficiency as novice individuals practiced the lifeboat task over a relatively short period. One of our hypotheses was that in the early trials when the subject is still new at the task, the patterns of neural activity may be very different from one trial to the next as the individual continuously adjusts his/her strategy based on their experience from the previous trial. Then, as the individual becomes more comfortable and proficient with the task and settles on an effective strategy for task execution, the pattern of neural activity may become more stable across consecutive trials. Borghini et. al. (2017) used a similar hypothesis in their work [39], but they tested this hypothesis over sessions/days. We wanted to test the same hypothesis over a shorter training period. To assess this, we used machine learning to classify 2-second epochs of data from adjacent trials and used the classification accuracy as a measure of data similarity (the lower the classification accuracy, the less separable is the data, and thus the more similar are the two trials).

Statistical classification was done on the data from adjacent trials for each participant individually, as well as with the data from all participants combined. Trial n vs. Trial n+1 classification was done where n varied from 1 to 9 (i.e., Trial 1 vs. 2, Trial 2 vs. 3 …. Trial 9 vs. 10). Note that in the case where all participant data was combined, data for each participant was first (i.e., before being combined into one dataset) normalized from 0 to 1 as follows:

$$Normalized\ feature\ value = \frac{feature\ value - \min(all\ samples)}{\max(all\ samples) - \min(all\ samples)} \qquad \text{(Equation 3.3)}$$

where "all samples" includes epochs from both trial n and trial (n+1). Again, this normalization was done for each frequency band/electrode combination.

This process of classifying adjacent baseline trials will be described in greater detail in the following sections.

3.8.1 Individual alpha frequency (IAF)

Though fixed frequency ranges are often used to define the different EEG signal bands (e.g., delta, alpha, theta, beta, and gamma), there is actually considerable variability in EEG among individuals, and it is often preferred to calculate individualized frequency ranges for these bands of interest. The individual alpha frequency (IAF) is one of the most common techniques. The IAF is associated with the maximum power of resting eyes-closed EEG rhythms [48]. In this experiment, we first determined the frequency of maximum power for each of the three eyes-closed baseline trials for a given individual using FFT-based power spectrum analysis (Welch technique). That individual's IAF was then calculated to be the mean of these three frequency values. With reference to the IAF, the bands of interest were then determined as follows: delta (IAF - 8 Hz to IAF - 6 Hz), theta (IAF - 6 Hz to IAF - 4 Hz) and alpha (IAF – 4 Hz to IAF + 2 Hz); fixed bands were defined for beta (13–30 Hz) and gamma (30–40 Hz) [48]. The mean IAF peak across participants was 9.7 Hz (±0.2 standard error of mean, SEM).

3.8.2 EEG power calculation

To calculate the average power, the time series EEG data was divided into 2 second epochs. The sampling frequency was 500Hz. So, there were 1000 samples in each epoch.

Then the average power of each frequency band was calculated for those 1000 samples using equation 3.4. The average power of a signal x(t) as a function of t is defined as:

$$P = \lim_{T \to \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} |x(t)|^2 dt \qquad \text{(Equation 3.4)}$$

Based on the literature, the delta, theta, and alpha were the frequency bands that seemed the most promising for identifying patterns related to training/learning, so only these three bands were considered in this analysis.

3.8.3 Electrode selection and total number of features

The EEG signal power in the three different frequency bands of interest (delta, theta, alpha) were calculated for individual electrodes. EEG signals were collected from a total of 64 electrodes, however all were not used in the analysis. Electrode FCz served as the reference electrode, and electrodes Tp9 and Tp10 were used to record EOG activity (which was note used in this analysis). Among the remaining 61 electrodes, nine (specifically Cz, C1, C2, C3, C4, C5, C6, T7 and T8) are associated with the motor cortex region (shown in Figure 3.8). Because the motor cortex is responsible for voluntary movement and is not directly related to learning, we eliminated those electrodes from the analysis.

Electrodes FT9, FT10, TP9, and TP10 were also excluded because for some participants we could not obtain good conductivity with the scalp, resulting in poor signal quality. A total of 50 electrodes were included in the analysis of which 24 were over the frontal region, and 26 were over the parietal and occipital regions. The name and position of all 50 selected electrodes are shown in the image below:
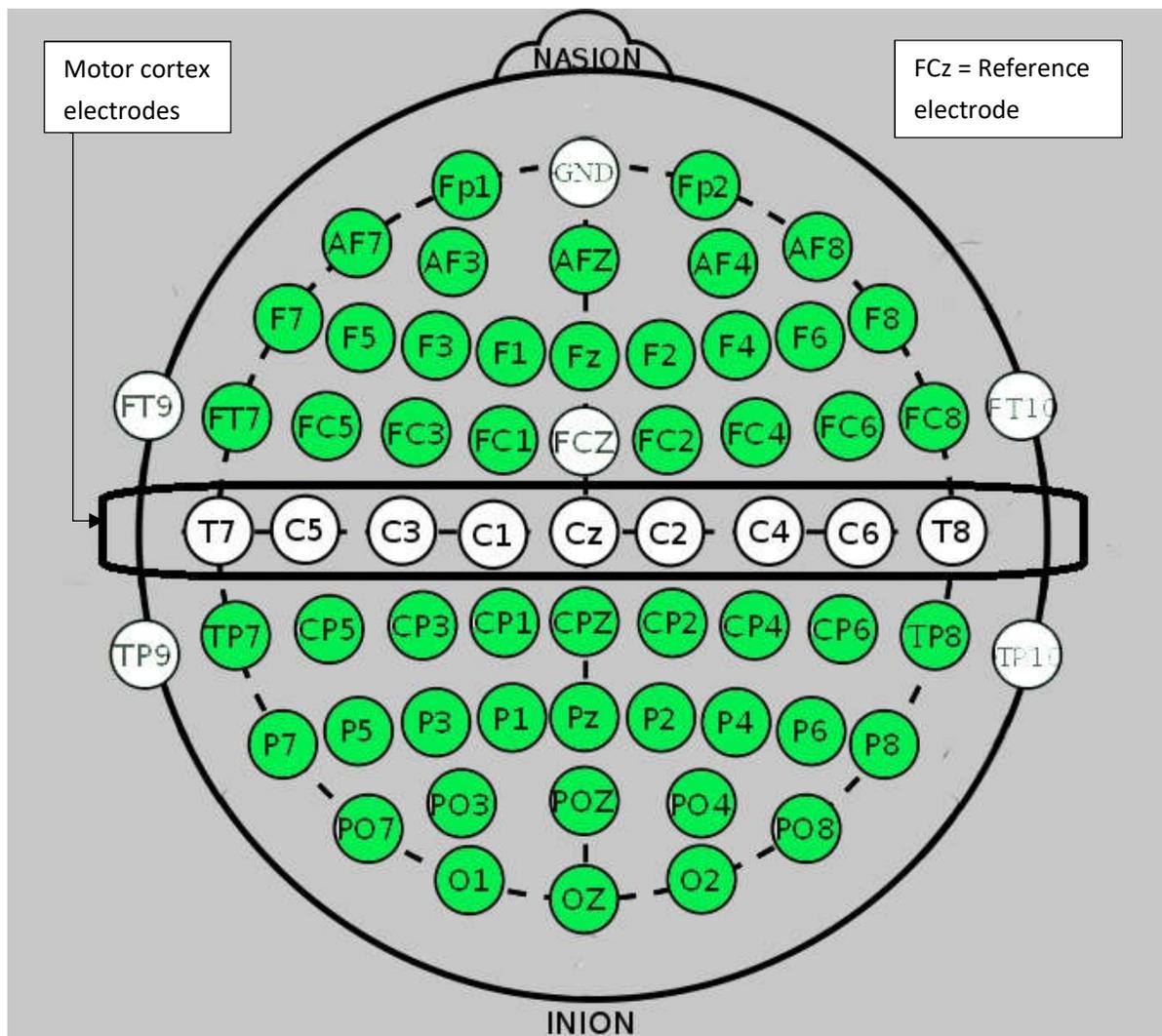
Figure 3. 8 - Electrodes used in the data analysis (shaded)

So, there were 50 electrodes included in the classification analysis, and for each one power was calculated in three different frequency bands. Thus, the total number of features was 3 x 50 = 150.

3.8.4 Fisher criterion for automatic feature selection

The characteristics of the best feature is that this feature should be quite distinguishable between the two classes, in this case between Trial n and Trial n+1. To evaluate how

discriminatory each feature was, the Fisher criterion was used. By selecting the features with the highest Fisher criterion value, you are left with features that maximize the distance between the means of the two classes while minimizing the variance within each class. The Fisher criterion, J, for a particular feature, w, is given by equation 3.5 below,

$$J(w) = \frac{|m_1 - m_2|^2}{s_1^2 + s_2^2}$$
(Equation 3.5)

Where m_x represents the mean of the samples (in this case EEG epochs) from class x, and s_x^2 represents the variance of the samples from class x

Thus, we calculated J between adjacent trials for each of the features and selected the three features with the highest J values.

3.8.5   Linear Discriminant Analysis (LDA)

The classification was done using a linear discriminant analysis based on Bayes' discriminant rule. A linear classifier in general tries to establish a hyperplane separating the signal space into individual subspaces for all classes. In the binary case, the decision rule for a given vector x to belong to class C_1 and not C_2 reads:

$$If\ p(C_1|x) > p(C_2|x)\ then\ x \in C_1\ and\ x \in C_2 otherwise$$

These probabilities can be computed using Bayes' formula:

$$p(C_K|x) = p(C_K)p(x|C_K)p(x)$$
(Equation 3.6)

Where $p(C_k)$ is the prior probability for a class k and $p(x|c_k)$ is the class distribution. Assuming that all classes are a priori equally probable, the priors can be neglected here. Therefore, the decision rule reduces to

$$p(C_1|x) > p(C_2|x)\ then\ x \in C_1\ and\ x \in C_2 otherwise$$

In the case of LDA, a multivariate Gaussian distribution is assumed for each of the classes'

C_k, given by

$$p(x|c_k) = \frac{1}{\sqrt{(2\pi)^f \det(c)}} \exp(-\frac{1}{2}(x - \mu_k)^T c^{-1}(x - \mu_k)) \qquad \text{(Equation 3.7)}$$

Where x is the vector to be classified, f is the dimension of this vector, c is the common covariance matrix for all classes, and μ_k is the mean value of class k.

3.8.6 Five-fold cross validation

Cross-validation is a model validation technique for assessing how the results of a statistical analysis will generalize to an independent data set [49]. The goal of cross validation is to define a dataset to "test" the model in the training phase and give an insight on how the model will generalize to an independent dataset.

One of the main reasons for using cross-validation instead of using the conventional validation (e.g., partitioning the data set into two sets of 70% for training and 30% for test) is that there is not enough data available to partition it into separate training and test sets without losing significant modelling or testing capability. In these cases, a fair way to properly estimate model prediction performance is to use cross-validation as a powerful general technique.

In k-fold cross-validation, the original dataset is randomly partitioned into k equal sized subsets. Of the k subsets, a single subset is retained as the test data for testing the model, and the remaining k − 1 subsets are used as training data. Feature selection and classification are performed using these training and test sets. This process is then repeated k times, with each of the k subsets being used exactly once as the test set.

Classification accuracy is calculated as the percentage of samples from the test set that were correctly classified by the model. The k classification accuracy results from the k folds can then be averaged to produce a single estimation. The advantage of this method over repeated random sub-sampling is that all observations are used for both training and validation, and each observation is used for validation exactly once. 10-fold cross-validation is commonly used [50], but in general k remains an unfixed parameter.

In this analysis k was taken to be five. To minimize the effect of random selection of data, the cross-validation procedure was repeated for 1000 iterations (i.e., the data was randomly divided into five datasets 1000 times) and then the average classification accuracy was calculated across all 1000 runs. Note that in each "fold" of all 1000 runs, only the training data was used for feature selection and classifier training, and only the testing data was used for classifier testing.

Figure 3.9 shows a sample procedure of one "fold" of five-fold cross validation for our dataset.
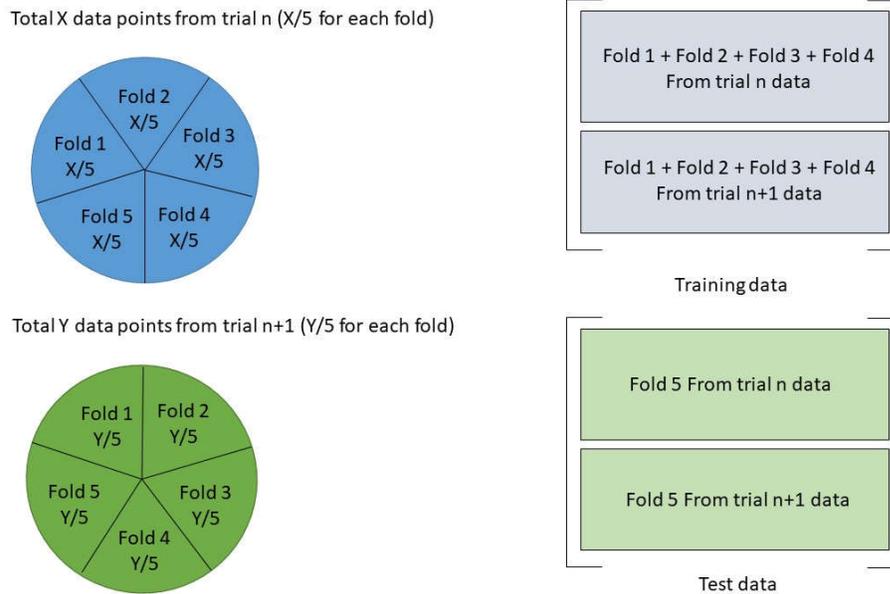
Figure 3. 9 – Example of data division for one "fold" of 5-fold cross validation

3.8.7   Sensitivity and specificity

In this experiment, lifeboat trial lengths were not of fixed time length. For that reason, the number of samples (i.e., epochs of EEG data) were not equal from trial to trial. In terms of the classification analysis, this resulted in the classes being "unbalanced". When classes are significantly unbalanced, the one class which is infrequently present is most likely to be predicted as rare occurrences, undiscovered or ignored, or assumed as noise or outliers which results in more misclassifications of the minority class compared to the prevalent class [51,52]. To account for the potential bias in the classification accuracy (i.e., percentage of correctly classified samples), we instead calculated the "adjusted accuracy" as follows [53]:

$$adusted\ accuracy = \frac{sensitivity + specificity}{2} \qquad \text{(Equation 3.8)}$$

where

$$Sensitivity = \frac{number\ of\ true\ positives}{number\ of\ true\ positives + number\ of\ false\ negatives} \quad \text{(Equation 3.9)}$$

$$Specificity = \frac{number\ of\ true\ negatives}{number\ of\ true\ begative = number\ of\ false\ positives} \quad \text{(Equation 3.10)}$$

and

true positive = Trial n sample classified correctly as Trial n

true negative = Trial n+1 sample classified correctly as Trial n+1

false positive = Trial n+1 sample classified incorrectly as Trial n

false negative = Trial n sample classified incorrectly as Trial n+1

Note that this terminology ("true/false positives/negatives") makes more sense intuitively when discussing such things as the diagnosis of diseases but can be applied to any binary classification problem by assigning (often arbitrarily) one of the classes as "positive" and one as "negative" (in the above description we have assigned "Trial n" to be positive while "Trial n+1" is negative).

3.9   Statistical analysis of trends in the data

The major statistical test that was performed here on the data is the statistical classification between adjacent trials for both individual subjects and mixed data of all subjects. Multiple linear regression method and paired t-test was used to test the significance of the trend in classification accuracy over trials. Correlation between several combinations of scores (NASA-TLX vs performance score, NASA-TLX vs classification accuracy and Performance score vs classification accuracy) were showed as well using the graphical

correlation method. At last the frequency of each feature getting selected for the classification was calculated and showed in a bar plot.

4   Chapter 4: Results

Of the 14 datasets collected, only 10 were included in the analysis. One participant performed so poorly that he was unable to finish the task and thus his data were excluded. For three other participants, the EEG data had to be excluded due to excessive motion artifact.

4.1   Task performance

As described in the previous chapter, Equation 3.1 was used to calculate each participant's performance score for each trial. The scores for each participant are given in Table 4.1, and the mean score across participants and trials is shown in Figure 4.1. In general, an upward trend can be observed in both the individual and mean scores. A paired t-test reveals a significant increase in performance score in Trial 10 as compared to Trial 1 (h=1,p = 0.012).

Table 4. 1 : Performance score for all participants

Trial #	1	2	3	4	5	6	7	8	9	10
Participant 1	0.20	0.26	0.42	0.50	0.39	0.34	0.45	0.47	0.53	0.39
Participant 2	0.19	0.30	0.31	0.29	0.30	0.28	0.31	0.31	0.33	0.33
Participant 3	0.39	0.41	0.38	0.45	0.49	0.46	0.46	0.51	0.46	0.54
Participant 4	0.44	0.25	0.44	0.47	0.49	0.59	0.46	0.46	0.42	0.64
Participant 5	0.23	0.36	0.51	0.50	0.55	0.58	0.59	0.58	0.51	0.54
Participant 6	0.26	0.22	0.23	0.18	0.34	0.35	0.32	0.30	0.33	0.31
Participant 7	0.37	0.37	0.42	0.42	0.45	0.42	0.46	0.43	0.51	0.43
Participant 8	0.46	0.46	0.42	0.44	0.37	0.35	0.39	0.37	0.37	0.38

Participant 9	0.40	0.40	0.37	0.36	0.34	0.34	0.35	0.34	0.34	0.37
Participant 10	0.38	0.30	0.25	0.32	0.39	0.48	0.46	0.45	0.27	0.56
Mean	0.33	0.33	0.37	0.39	0.41	0.42	0.43	0.42	0.41	0.45

For all participants, individual performance measures used in the calculation of the performance score are given below (Table 4.2). From this table it is clear that as the experimental session progresses, distance travelled per trial decreases, accuracy score increases and eventually plateaus at 25, and time taken also decrease. And all of these contributes to the increase in performance score.

Table 4. 2 : All performance measures for all participants

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7	Trial 8	Trial 9	Trial 10
	Participant 1									
Distance (m)	274.3	287.1	254.6	252.1	265	264.9	253.5	253.4	257.5	255.3
Accuracy score	13	18	25	20	23	23	25	25	25	25
Time (s)	163	218	217	160	223	233	217	211	182	211
Hits (hard/soft)	1/0	0/1	0/1	0/0	0/0	0/1	0/0	0/0	0/0	0/1
	Participant 2									
Distance (m)	285.9	268.7	264.3	270.3	262.7	265.8	264.7	261.4	264.5	263.9
Accuracy score	18	25	25	25	25	23	25	25	25	25
Time (s)	334	313	306	315	304	312	305	305	286	290
Hits (hard/soft)	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7	Trial 8	Trial 9	Trial 10
Participant 3										
Distance (m)	263.4	257.0	265.8	260.5	262.3	263.4	264.5	259.0	259.04	256.1
Accuracy score	21	25	20	25	25	25	25	25	25	25
Time (s)	207	235	197	214	196	206	205	191	209	180
Hits (hard/soft)	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Participant 4										
Distance (m)	259.6	343.8	279.7	253.8	256.6	265.6	259.5	255.2	298.5	253.4
Accuracy score	25	20	25	25	25	25	25	25	25	25
Time (s)	182	188	163	209	199	160	208	178	200	215
Hits (hard/soft)	1/0	1/0	0/0	0/0	0/0	0/0	0/0	1/0	0/0	0/0
Participant 5										
Distance (m)	274.3	267.1	262.1	257.7	260.3	262.1	260.7	259.9	261.9	258.1
Accuracy score	13	23	25	23	25	25	25	25	23	25
Time (s)	179	196	158	179	162	164	162	167	172	180
Hits (hard/soft)	0/1	1/0	1/0	0/0	0/1	0/0	0/0	0/0	0/0	0/0
Participant 6										
Distance (m)	264.5	257.5	263.1	278.9	255.3	265.5	255.3	264.6	253.7	261.2
Accuracy score	19	19	19	23	23	25	25	25	25	25
Time (s)	281	300	314	379	262	266	302	316	297	308
Hits (hard/soft)	0/0	0/1	0/0	1/0	0/0	0/0	0/0	0/0	0/0	0/0

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7	Trial 8	Trial 9	Trial 10
Participant 7										
Distance (m)	258.1	256.8	253.9	253.7	251.9	256.3	266.6	262.7	256.9	254.2
Accuracy score	25	25	25	25	25	25	25	25	25	25
Time (s)	261	262	233	237	205	213	206	216	190	209
Hits (hard/soft)	0/0	0/0	0/0	0/0	0/1	0/1	0/0	0/0	0/0	0/1
Participant 8										
Distance (m)	253.9	266.8	257.3	265.6	253.7	256.7	256.3	254.3	255.1	263.5
Accuracy score	25	25	25	23	25	25	25	25	25	25
Time (s)	216	204	230	196	267	279	251	242	274	247
Hits (hard/soft)	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/1	0/0	0/0
Participant 9										
Distance (m)	256.1	253.1	255.6	253.5	255.3	251.4	252.9	250.5	252.1	251.8
Accuracy score	25	25	25	25	25	25	25	25	25	25
Time (s)	227	248	266	276	289	292	285	290	290	270
Hits (hard/soft)	0/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Participant 10										
Distance (m)	266.7	265.5	292.4	273.3	267.1	263.1	265.6	253.5	278.5	259.8
Accuracy score	25	21	23	23	25	25	25	25	23	25
Time (s)	246	260	283	262	240	200	206	211	301	173
Hits (hard/soft)	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
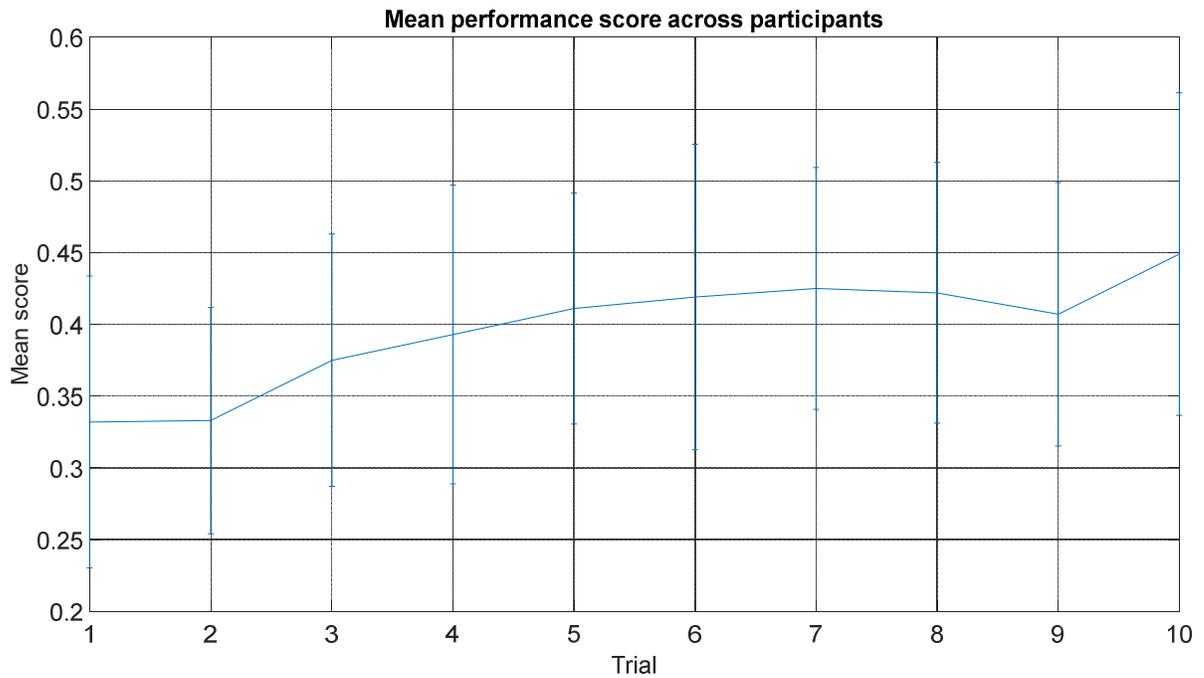
Figure 4. 1 - Mean performance score across all participants with error bars representing standard deviation

4.2   NASA-TLX score

The mental workload scores were calculated for each participant and each trial based on their responses to the NASA-TLX questionnaire (see Figure 4.2). The subjective workload scores for all participants over the ten trials are shown in Table 4.3, and the mean score across participants can be seen in Figure 4.2. In general, a downward trend can be observed in both the individual and mean scores. A paired t-test reveals a significant decrease in perceived workload rating in Trial 10 as compared to Trial 1 (h=1, p = 0.016).
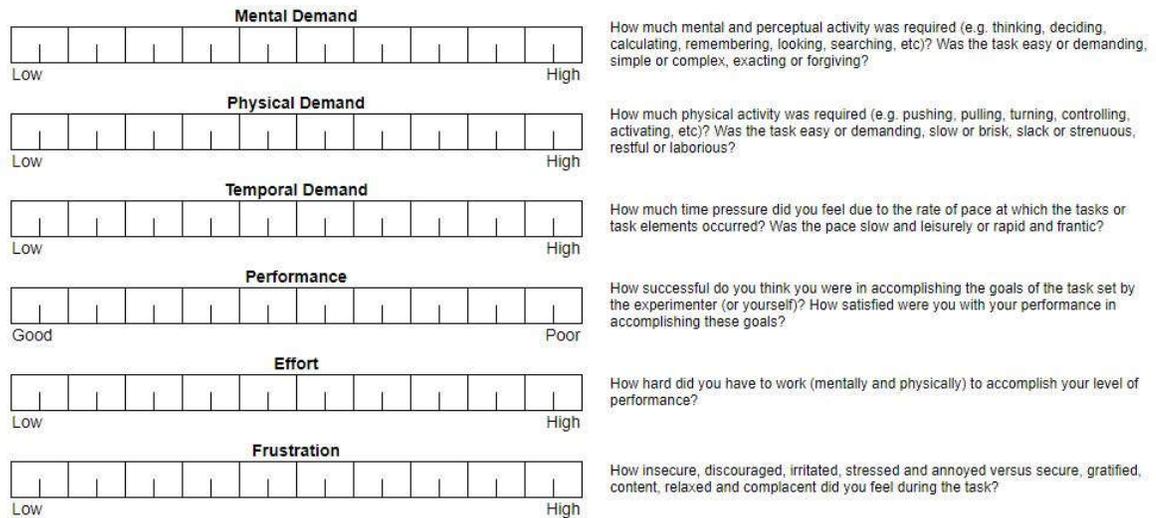
Figure 4. 2 : NASA-TLX questionnaire

Table 4. 3 : NASA-TLX score for all participants

Trial #	1	2	3	4	5	6	7	8	9	10
Participant 1	63.33	57.50	51.67	48.33	48.33	41.67	53.34	46.67	44.17	47.50
Participant 2	45.00	43.33	40.83	39.16	39.16	47.50	41.66	45.00	45.00	45.83
Participant 3	51.66	55.00	70.00	40.83	16.66	26.66	22.50	10.00	11.66	18.33
Participant 4	69.16	93.33	88.33	60.83	55.83	53.33	42.50	40.00	36.66	15.83
Participant 5	57.50	44.16	34.16	34.16	33.33	33.33	30.00	30.83	28.33	36.66
Participant 6	45.00	69.16	76.67	64.16	45.00	55.00	42.50	54.16	34.16	50.00
Participant 7	56.66	56.66	49.16	50.00	48.33	54.16	52.50	53.33	49.16	54.16
Participant 8	50.00	40.00	33.33	38.33	24.16	20.83	24.16	22.50	22.50	16.66
Participant 9	66.66	60.00	61.67	62.50	59.16	58.33	55.00	55.00	54.16	55.83
Participant 10	53.33	53.33	58.33	56.67	50.83	48.33	48.33	42.50	58.33	46.66
Mean	55.83	57.25	56.41	49.50	42.08	43.91	41.25	40.00	38.41	38.75
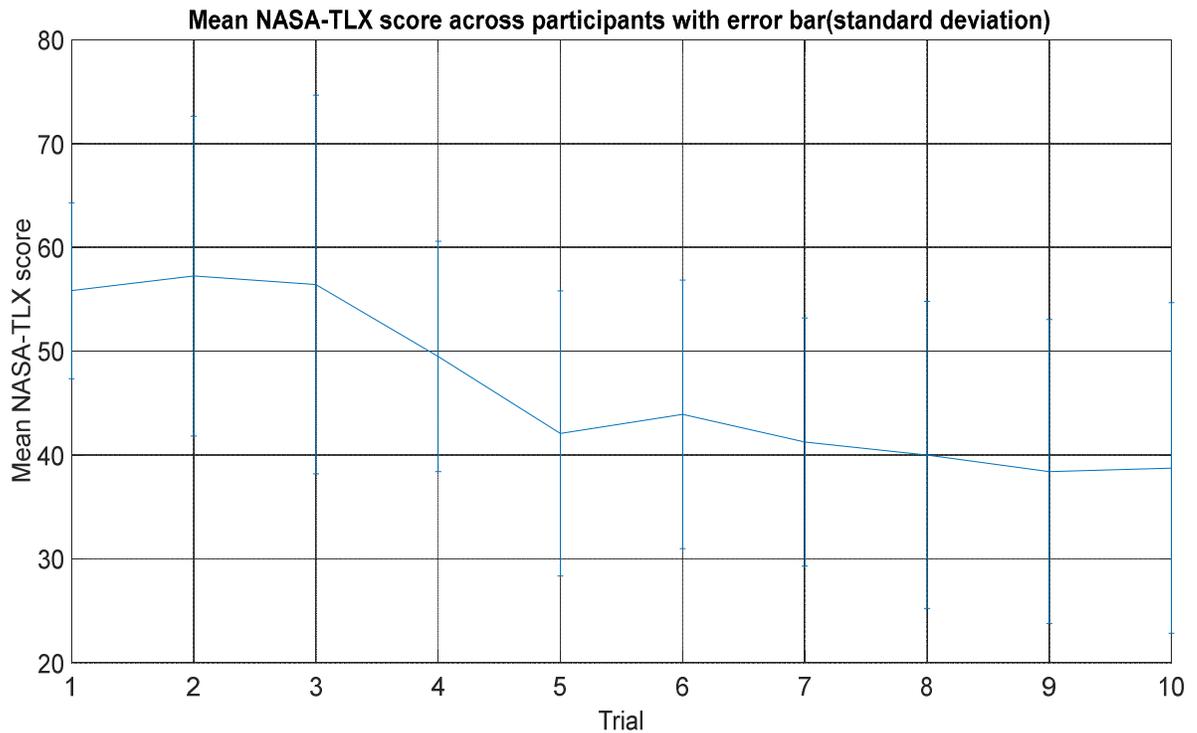
Figure 4. 3 – Mean NASA-TLX score with standard deviation

4.3 EEG data

For EMG and other types of noise (e.g., swaying and swinging) approximately 17% of our total data points have been discarded. For any given participant, this rate is highest for 3rd participant where 30% of his total data was discarded for EMG and other types of noise.

Across all the participants, the mean duration for a single trial was 198.4 ± 60.81 seconds (mean ± std). The duration for the shortest trial was 1 minutes and 02 seconds (Participant #3, Trial #1) and the longest trial was 5 minutes and 22 seconds long (Participant #6, Trial #4).

4.3.1   Adjacent trial classification accuracy

The classification accuracies of adjacent trials for each participant, along with the mean accuracies across participants, are given in Table 4.4, and the mean is plotted in Figure 4.4.
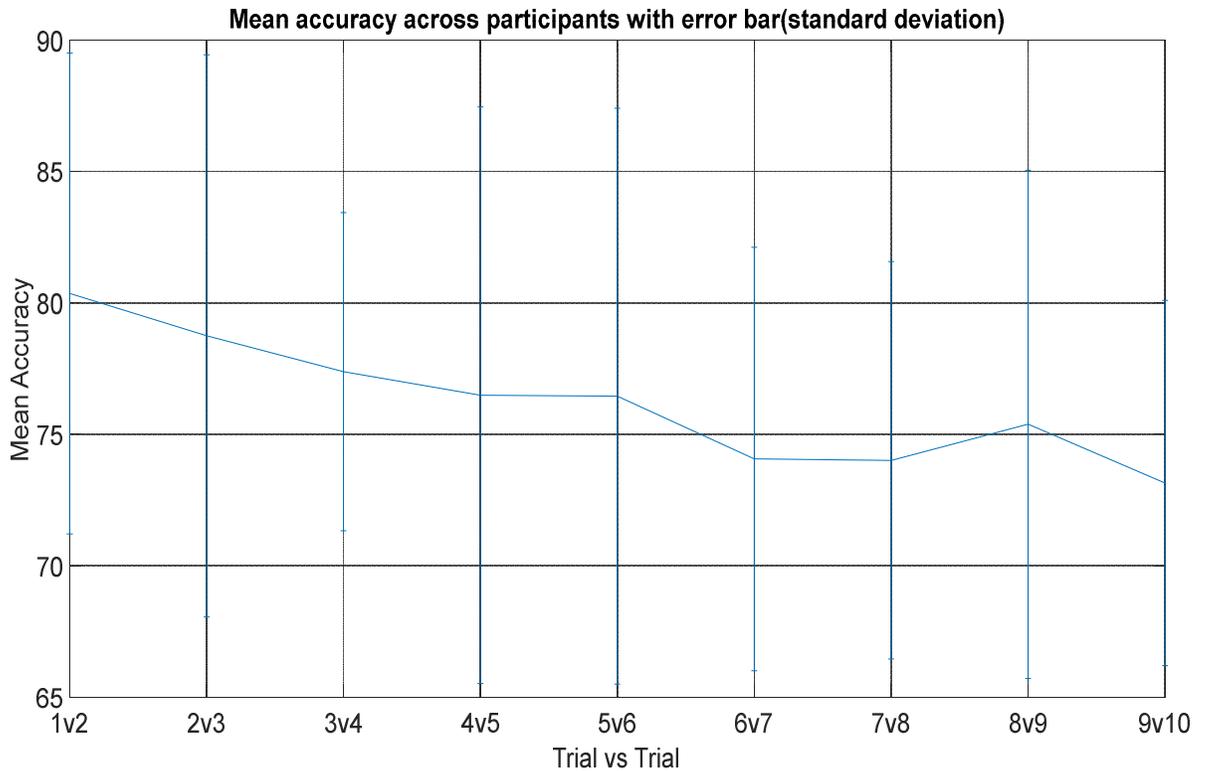
Figure 4. 4 – Mean adjusted classification accuracy for adjacent trials with error bars representing standard deviation

While there is significant variation in the trend for individual subjects, the average trend is clearly downward, with the average accuracy decreasing from around 80% to 73%. Multiple linear regression confirms a statistically significant downward trend in adjacent trial classification accuracies over the course of the session ($p < 0.001$, DF=79, F value = 13.57). Furthermore, a paired t-test indicates that the average classification accuracy for

Trial 1 vs. Trial 2 is significantly different (higher) than that for Trial 9 vs. Trial 10 (h=1, p=0.019).

Table 4. 4 : Classification accuracy between adjacent trials for all participants[1]

Trials classified	1 vs 2	2 vs 3	3 vs 4	4 vs 5	5 vs 6	6 vs 7	7 vs 8	8 vs 9	9 vs 10
Participant 1	87.7	82.8	82.8	66.4	67.7	68.2	77.2	77.3	74.8
Participant 2	94.7	97.7	73.0	70.0	85.9	81.1	87.9	90.6	78.3
Participant 3	76.3	74.1	82.2	69.9	78.5	78.9	68.4	64.9	84.2
Participant 4	89.4	87.6	84.3	99.2	94.4	76.9	81.5	91.2	81.4
Participant 5	89.7	92.5	80.3	88.2	90.7	86.9	78.6	73.6	74.8
Participant 6	67.5	68.8	74.3	77.5	62.1	59.7	68.3	69.5	68.7
Participant 7	76.3	72.6	78.7	83.6	73.7	76.1	77.2	81.3	67.6
Participant 8	72.1	73.8	65.2	65.1	64.1	66.8	68.3	69.6	70.4
Participant 9	74.8	70.1	80.9	76.1	72.9	77.3	67.0	64.6	61.2
Participant 10	74.8	67.2	71.9	68.6	74.3	68.4	65.3	70.8	69.7
Mean	**80.4**	**78.8**	**77.4**	**76.5**	**76.5**	**74.1**	**74.0**	**75.4**	**73.1**

[1] If we consider the number of epochs for all participants and all trials, the lowest was 31 (most were much higher, 98 ± 29 (average $\pm$ std); see Appendix B). The threshold limit for statistical significance for a 2-class problem at this worst-case scenario of 30 samples per class at α=5% is 62.5% accuracy [35]. As can be seen from this table, most of the accuracies from all participants are greater than this value. Participant 6 and 9 have some values that are less than this one. But according to their individual lowest number of epochs these accuracies are significant as well (>57.5% accuracy). Therefore, we can say that all accuracies are greater than random chance (i.e, 50%).

A similar downward trend is observed for adjacent trial classification accuracies determined using all the participant data combined (see Figure 4.5). Here the classification accuracies are generally much lower, starting from around 60% for Trial 1 vs. Trial 2 and plateauing at around 50% by Trial 5.
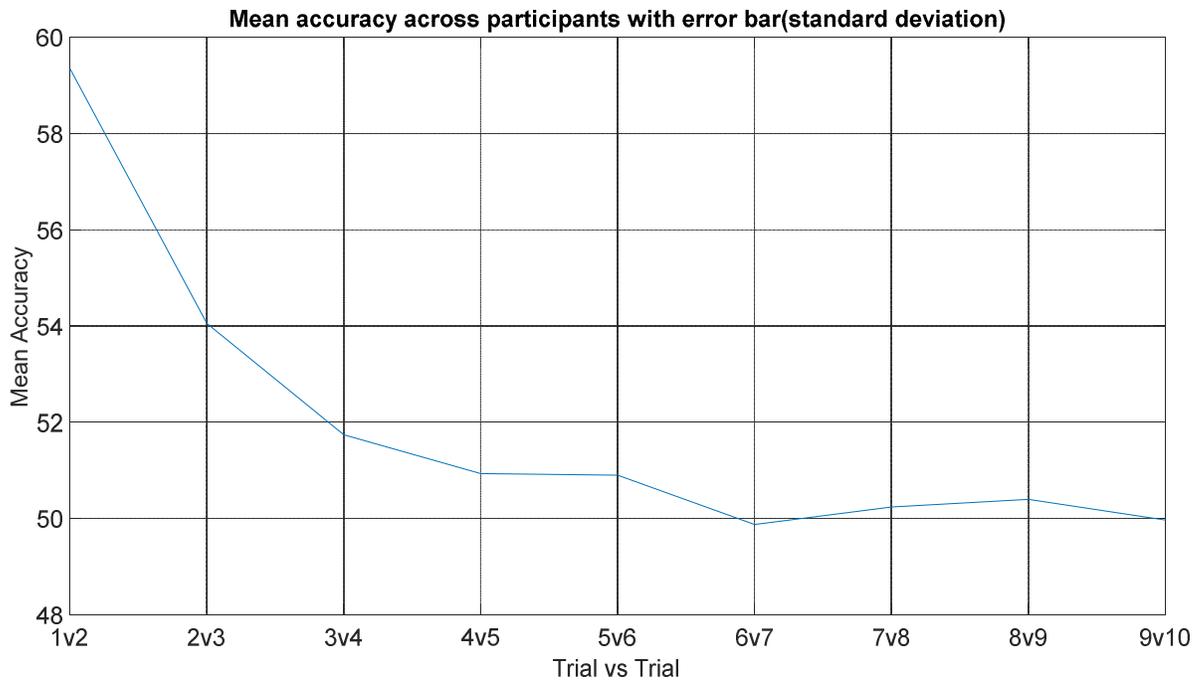
Figure 4. 5 - Adjacent trial classification accuracies for all participants' data combined

4.4   Most relevant features

The frequency with which individual features were selected by the automatic feature selection algorithm was investigated in order to determine which electrodes and frequency bands contributed most significantly to the task-related brain activity. Three features were selected for each fold of each run of the cross-validation analysis. This is a total of 5 x 1000 = 5000 selections. Figure 4.6 shows feature selection frequency for the combined data of all participants.
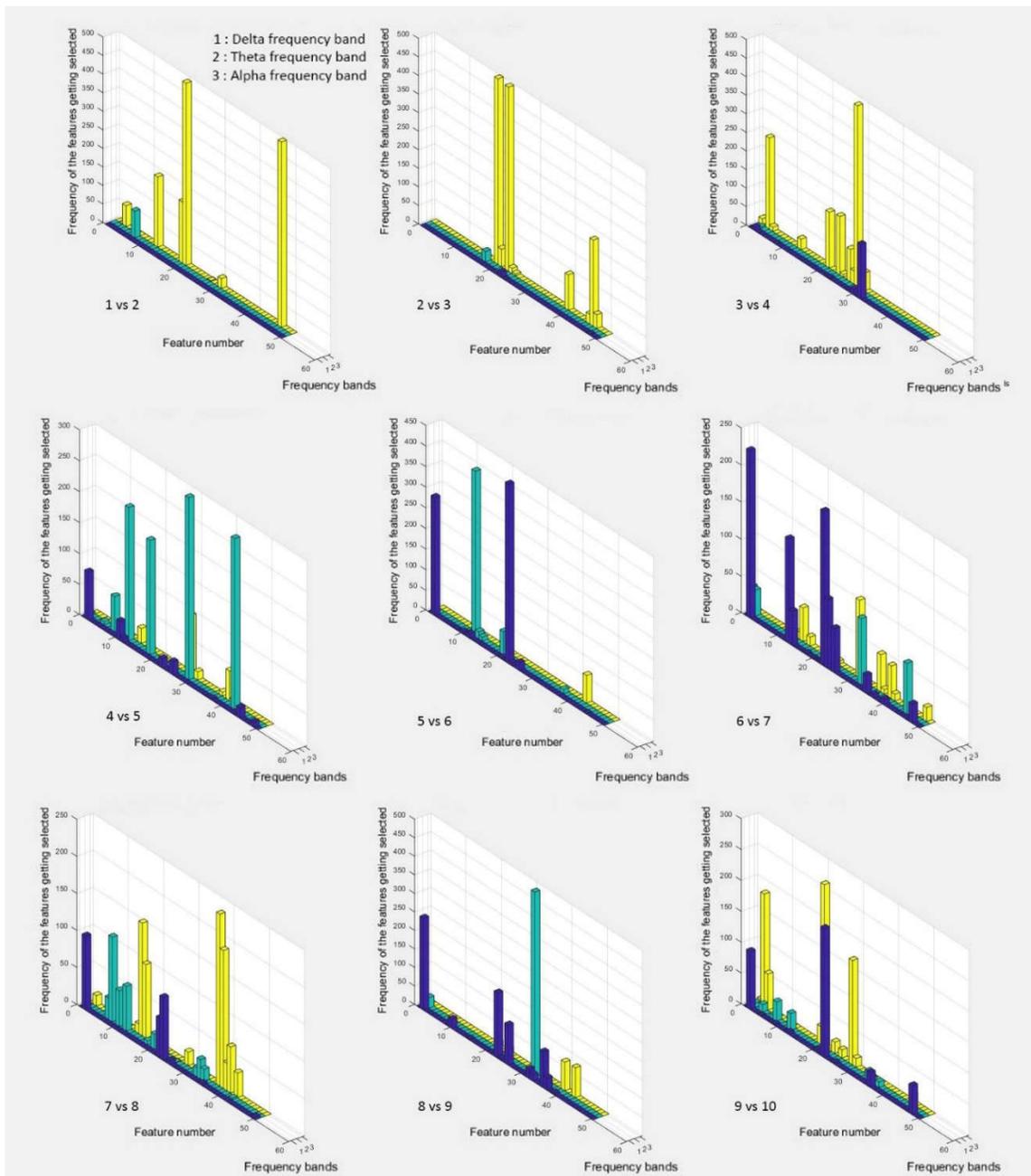
Figure 4. 6 : Features being selected for adjacent trials

It can be observed from these plots that the features most often selected by the feature selection algorithm in the early trials of the session are more likely to be from the alpha

frequency band, whereas for the later trials the frequency of the features being selected are more evenly distributed and include the theta and delta bands as well.

4.4.1 Frontal theta, parietal alpha and overall beta decrease

For each participant, alpha power was calculated over the parietal region (mean of electrodes Pz, P1, P2, P3, P4, P5, P6), the theta power was calculated over the frontal region (mean of electrodes Fz, F1, F2, F3, F4, F5, F6), and beta power was calculated over all regions (mean of all electrodes) for each trial. Figure 4.7 shows the group mean trends across all ten trials. There is a slight downward trend in all three frequency bands/regions, and this trend is significant for parietal alpha and overall beta according to multiple linear regression (parietal alpha, $p<0.001$, DF $= 89$, F value $= 12.75$; frontal theta, $p = 0.281$, DF$=89$, F value$=4.99$; overall beta $= 0.005$, DF$=89$, F value$=8.44$). However, no significance was found via paired t-test between Trial 1 and Trial 10 (parietal alpha, $p=0.827$; frontal theta, $p = 0.668$; overall beta, $p = 0.339$).
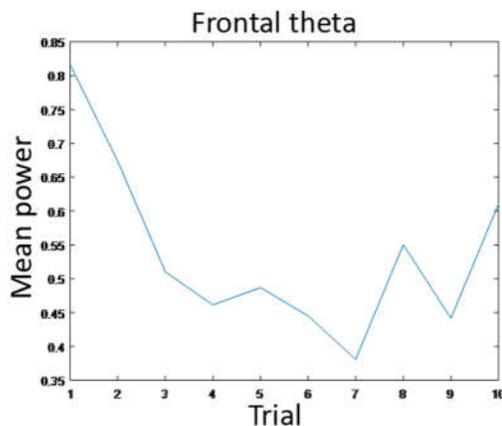
Figure 4. 7 a - Mean frontal theta power   Figure 4.7 b - Mean parietal alpha power
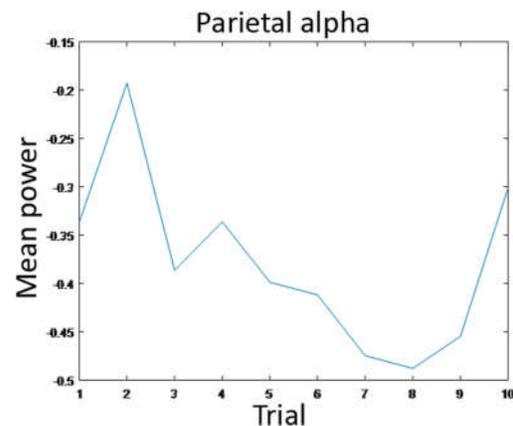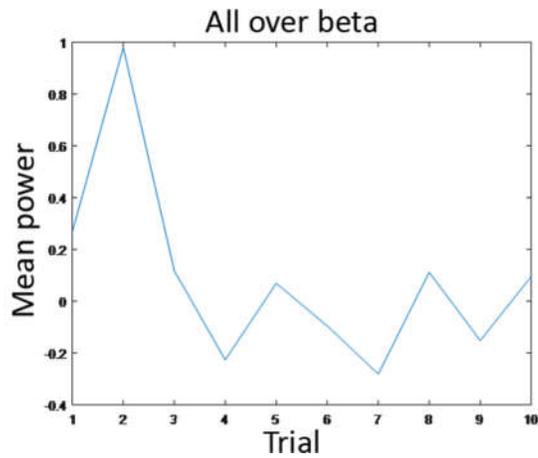
Figure 4.7 c - Mean beta power

4.5   Correlation between NASA-TLX score and performance score

The correlation between the perceived workload rating and performance score was investigated using the graphical correlation method (see Figure 4.8).

$$y = -0.0049x + 0.6221$$
$$R^2 = 0.8406$$
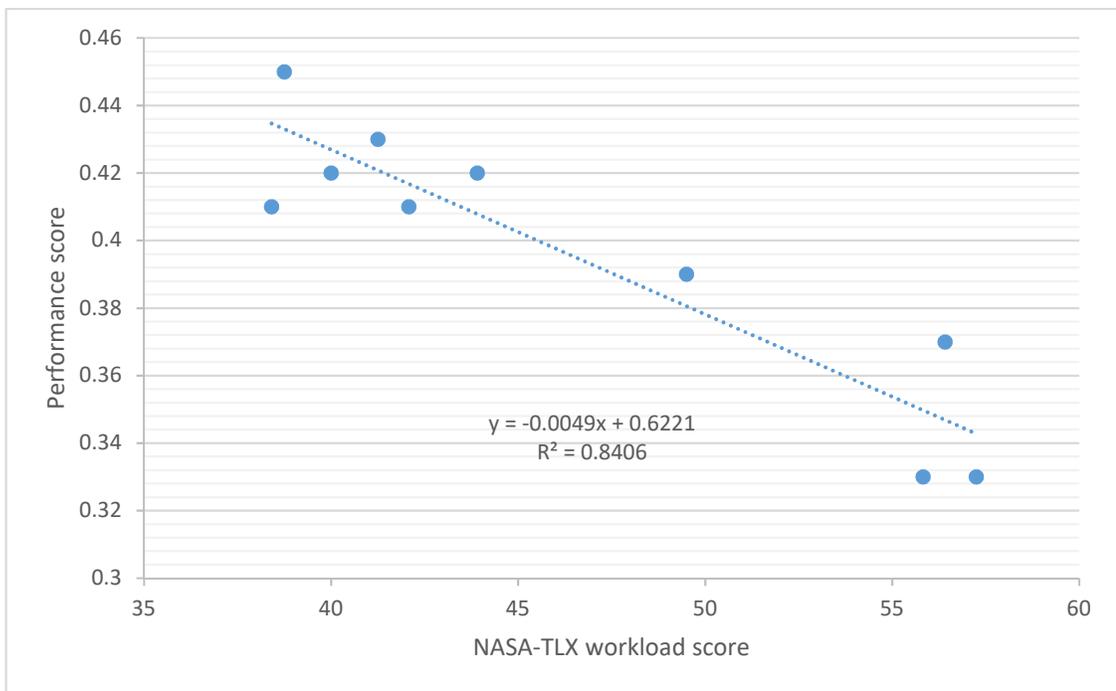
Figure 4. 8 - Correlation between NASA-TLX score and performance score

Here, this figure shows the scatter plot between mean-performance and NASA-TLX workload score. The R^2 value is 0.8406 which means that the regression model is a very good fit for the data. The regression analysis reported significant (p<0.001) and high correlation (|R|=0.9168) between the measures. As we would expect, as the task performance score increases, the perceived mental workload score decreases.

4.6   Correlation between NASA-TLX and classification accuracy

When novice participants perform a task for the first time, they use all their cognitive resources to perform the task resulting in a wide variety of activity throughout the whole brain. But later as they get used to the task, their use of cognitive resources becomes more limited, and they do not have to think so much about performing the task [54]. A trained person efficiently attends to information that is relevant to task performance, while leaving the irrelevant and potentially distracting information unattended.

Again, from Figure 4.9, we can say that NASA-TLX score and classification accuracy are positively correlated, which means as the NASA-TLX score decreases, classification accuracy follows. And NASA-TLX score is likely to decrease as an individual becomes more confident in performing a task.

NASA-TLX vs Classification accuracy

$y = 0.2962x + 62.81$
$R^2 = 0.8453$
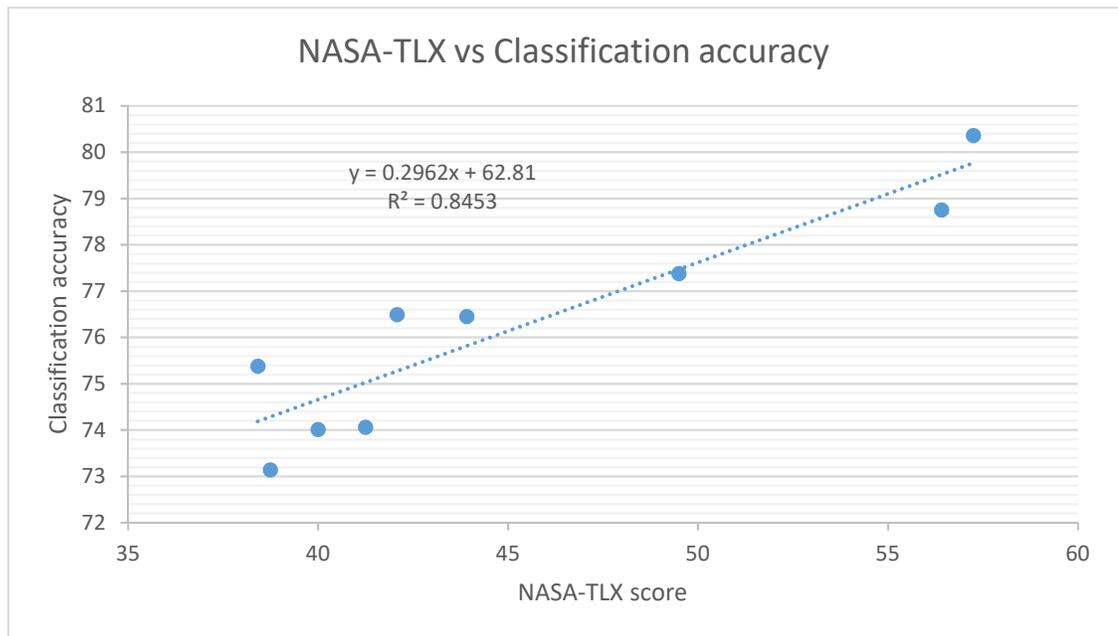
Figure 4. 9 - NASA-TLX vs classification accuracy

4.7 Correlation between performance score and classification accuracy

On the other hand, task performance score and classification accuracy show negative correlation to each other (Figure 4.10). As the task performance score increases, the classification accuracy between consecutive trials decreases.
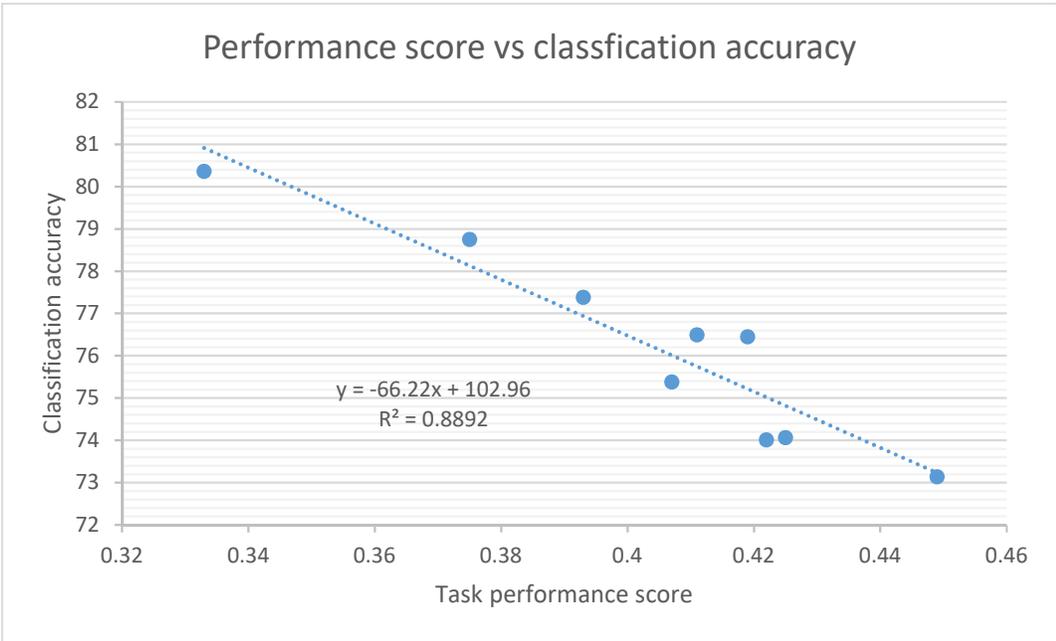
Figure 4. 10 - Performance score vs classification accuracy

5   Chapter 5 : Discussion

Passive brain–computer interfaces provide information regarding a user's mental state to a computerized application for the purpose of enhancing the interaction with the application. The main objective of this experiment was to work toward the development of a passive BCI for assessing skill acquisition of participants undergoing virtual environment (VE)-based training. One of the current limitations of standard training assessment procedures, which are based primarily on performance measures, is indeed the lack of objective information about the amount of cognitive resources required by the trainees during the operative activity. So, in this research, trends in neural activity was explored as participants practiced a task over a relatively short, single session training period that could potentially form the basis of an objective, cognitive measure of skill acquisition that could be used in the development of a passive BCI for enhancing VE-based training programs.

5.1   Task-related neural activity becomes more stable with practice

It was hypothesized that in the early stages of training, when an individual is unskilled in the task, the patterns of neural activity may be very different from one trial to the next as the individual continuously adjusts his/her strategy based on their experience from the previous trial. Then, as the individual becomes more comfortable and proficient with the task and settles into an effective strategy for task execution, the pattern of neural activity may become more stable across consecutive trials. For this analysis, the normalized EEG data was taken and performed classification on them on a trial vs. trial basis, for pairs of

adjacent trials (i.e, Trial #1 vs. #2, Trial #2 vs. Trial #3, etc.). Classification accuracy was assumed to be a measure of the "similarity" between the trials classified in terms of neural activity patterns (low accuracy would indicate higher similarity). On average across participants, a decreasing trend in the classification accuracies of adjacent trials was found over the course of the training. This indicates that as the participants got more practice with the task (i.e., completed more trials), the neural activity eventually started to get more stable, or similar, from trial to trial. This finding supports the above stated hypothesis. It can be observed from Table 4.4 that even though the decreasing trend in the mean classification accuracy over participants is quite significant according to multiple linear regression, the downward trend is not so obvious for individual participants. The classification accuracy between initial trials (Trial #1 vs. Trial #2, Trial #2 vs. Trial #3) are significantly higher than of later trials (Trial #8 vs. Trial #9, Trial #9 vs Trial #10) for most of the participants, but it followed a fairly random trend in the middle trials for many participants.

The variation in the middle trials might be due to the fact that people initially try to determine the effects of different ways they approach driving the lifeboat. During the first several trials they may try to adjust their strategies as they figure out what worked and what didn't, and this process would be different for everyone. However, according to performance measures and rating of perceived workload via the NASA-TLX questionnaire, all participants seemed to reach some level of competence with the task, and the brain activity did seem to at least begin to stabilize by the end, as indicated by the

lower classification accuracy of adjacent trials near the end of the session (i.e., Trial #9 vs. Trial #10) for most participants.

5.2 Improvement of performance score over the trials

Equation 3.1 was used to calculate the performance score for all the participants for all trials. There was no highest limit or lowest limit for this score. The equation was derived in such a way that the greater the distance covered, and time taken to complete the trial, the lower the score will be. The accuracy portion of the performance score was based on whether or not the participant hit any of the five buoys, and if so, the speed at which they hit them. Six and three points were deducted for a "hard" and "soft" hit, respectively. The overall performance scores for the consecutive trials followed an increasing linear trend, indicating that participants did improve over the course of the session. If we plot only the mean accuracy score for all the participants, we can clearly see that the score increases linearly from around 20 to 24.5 and then plateaus around 25 after 5th trial.
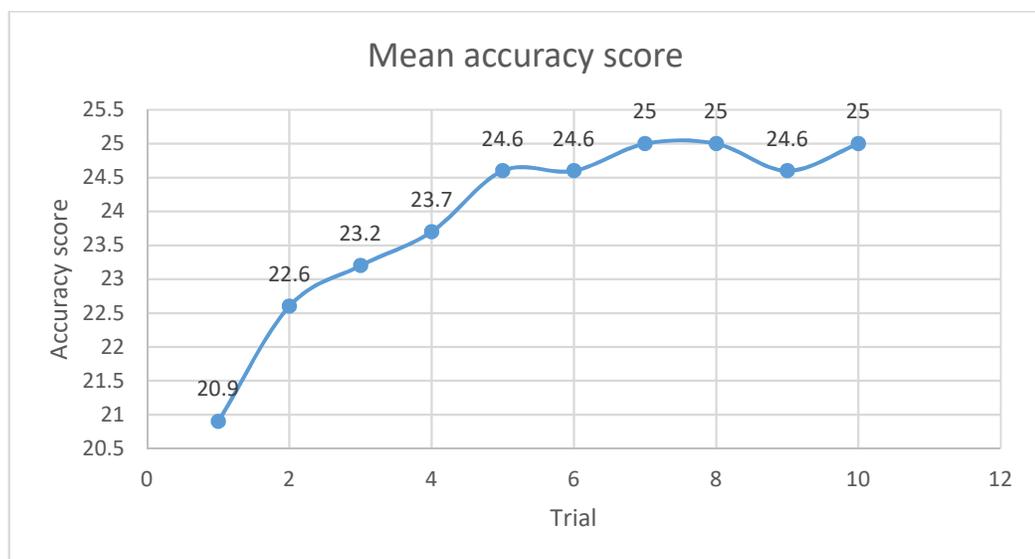
Figure 5. 1 - Mean accuracy score across participants

5.3   Verifying result with balanced classes

Generally, the length of time taken to complete the trials decreased as the participants became more skilled in the task. Therefore, later trials tended to be somewhat shorter than earlier trials, meaning that the amount of data used in the classification of adjacent trials earlier in the session was generally larger than for later in the session. According to literature, the larger the training set is, the more information it contains, and the more accurate the learned classifier can be (although there are other factors involved) [55]. It could happen that the decreasing trend observed in the classification accuracies was not due to relevant training-induced chances in neural activity, but rather was simply a result of there being less data for performing the classification analysis in the later trials.

To investigate if the reduced trial durations may have contributed to the observed decrease in classification accuracy, classification analysis was performed as described in Section 3.8, but with the classes balanced. To do this, for each participant I selected the shortest trial (i.e., with of the fewest epochs/samples), and randomly extracted the same number of samples from all remaining trials to use for classification. The resulting trend in classification accuracy was very similar in this case compared to when all the data was used, and there was still a statistically significant decreasing trend (p < 0.001, DF=79). This indicates that the observed trend was indeed due to changes in the neural activity and was not an effect of the difference in the trial lengths across the session.

Furthermore, if we consider the length of datasets, for first five trials the average number of samples (epochs) across participants are 184, 162, 148, 143 and 152 and for the later five trails they are 166, 136, 137, 141 and 207. The average number of samples goes up

and down but doesn't necessarily follow a decreasing trend over the trials. According to repeated measure one-way ANOVA, there is not a significant trend (p value: 0.989, DF = 89). So, we can say with confidence that the decreasing trend in the classification accuracy is not due to a decreasing number of samples across trials. Note that these are the average number of epochs after the EMG-related artifact removal process.

5.4 Decreasing power in several frequency bands

A deceasing trend in EEG power over multiple frequency bands and in both the frontal and parietal brain region was observed. Frontal theta and parietal alpha showed a decreasing trend of power as the trial increases (Figure 4.7a and 4.7b). The overall beta power also decreased. The downward trend parietal alpha and all over beta are significant according to multiple linear regression even though all three of these came out as not significant according to paired t-test.

A number of studies in the literature found that as an individual becomes familiar with a task, their frontal theta decreases and overall beta decreases [32, 4]. However, in this work, both frontal theta and parietal alpha followed roughly a U-shaped trend, which means they decreased in the middle trials and then increased again. A possible explanation for this is that in the beginning, participants usually have no idea how to perform the task, but by trial 5 or 6 most had improved quite significantly which may result in a decreasing power for both frontal theta and parietal alpha. At this point, participants may have started trying new ways (e.g. varying speed while turning, trying to complete the trial faster or slower) to perform the task in an attempt to further improve their performance. This type of strategy change was reported informally by several participants. This phenomenon may

be is responsible for the increasing trend in the later trials for both frontal theta and parietal alpha among which only parietal alpha was significant. Over beta power, however, started decreasing significantly starting from Trial 3, which is consistent with the trend reported in the literature.

5.5   Decreasing trend in NASA-TLX score

Table 4.3 shows that the mean NASA-TLX score varied for the participants starting from 55 to the lowest limit of 38. For all the participants, the score tended to decrease as the number of trials increased, indicating a reduction in perceived mental workload with task practice. This can possibly be the indication that as the participants became better at the lifeboat task, they need to think less about the execution of the task itself, and thus their workload decreases. While it's a subjective measurement and the scale is clearly different from person to person, the general trend is clearly decreasing across the session.

5.6   Limitations of this study

There are were some limitations of this study that are worth noting and addressing in future work.

First of all, data from only 10 participants were included in the analysis. Future work should involve verifying the results of this study with a larger dataset.

Also, only the power spectral features of the EEG signals (i.e., delta, theta and alpha frequency bands) were considered in the analysis. Though the literature supports the use of these measures, there are other features, for example relating to functional connectivity and network efficiency, that could be useful for identifying training-related trends in

neural activity. Future work should involve an investigation of a wider variety of EEG signal features.

Moreover, the performance score equation (Equation 3.1) was derived without any valid derivation metric. It was derived in such a way that if the accuracy score increases, the performance score increases. On the other hand, if the time taken and distance increases, performance score decreases. Weights for hard and soft hit was defined arbitrary as well even though hard hit was given double weight than soft hit. Though we are confident the correct constituent measures were included (e.g., hits, time, distance, etc.), how they were put together into a single metric may not have been ideal. This may have affected our ability to get an accurate trend in performance score.

Also, physiological factors such as boredom, engagement, attention, fatigue, habituation etc. could potentially have influenced the neural activity and contribute to the changes in classification accuracy. To mitigate the effects of this cofounding factors, we took baseline signals prior to each trial and normalized the trial data with the baseline data. Participants were also given necessary breaks (also water and food if needed) between the trials. We are therefore confident that the trend we observed was in fact due to training and not these other factors.

6   Chapter 6 : Conclusion

This thesis presented an analysis of changes of human neural activity patterns overtime as naive participants practiced a motor cognitive task in a virtual environment-based training simulator. Even though our long-term research objective was to develop a passive BCI to provide an objective, cognitive-based measure of task proficiency/learning, the short-term objective of this thesis was to identify EEG-based neural indicator(s) of task proficiency/learning over a short period of practice of a cognitive motor task performed in a VE-based training simulator.

The most significant result of this work was the identification of a training-related trend based on the classification accuracy between adjacent practice trials. On average, the classification accuracy significantly decreased as the participants became more skilled in the motor cognitive task, indicating that the neural activity begins to stabilize.

6.1   Future work

Though all participants clearly got significantly better at the lifeboat task over the course of the ten trials, data indicates that for most participants neither their performance nor their neural activity patterns reached a plateau in this period. Future work should involve training over a longer period of time (or in a simpler task that participants can reach competence in more quickly), so that the trend in neural activity can be observed beyond the point that performance plateaus to see if it provides any additional information about

task proficiency (i.e., do the neural signals continue to change after performance plateaus?).

Bibliography

[1] Zander TO, Kothe C. Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. J Neural Eng.

[2] F. Nijboer, F. O. Morin, S. P. Carmien, R. A. Koene, E. Leon, and U. Hoffmann, "Affective brain-computer interfaces: Psychophysiological markers of emotion in healthy persons and in persons with amyotrophic lateral sclerosis," 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, 2009.

[3] G. Borghini, P. Arico, F. Ferri, I. Graziani, S. Pozzi, L. Napoletano, J. P. Imbert, G. Granger, R. Benhacene, and F. Babiloni, "A neurophysiological training evaluation metric for Air Traffic Management," 2014

[4] G. Borghini, P. Arico, L. Astolfi, J. Toppi, F. Cincotti, D. Mattia, P. Cherubino, G. Vecchiato, A. G. Maglione, I. Graziani, and F. Babiloni, "Frontal EEG theta changes assess the training improvements of novices in flight simulation tasks," 2013

[5] Prataksita N, Lin Y-T, Chou H-C, Kuo C-H. Brain-robot control interface: Development and application. In: Bioelectronics and Bioinformatics (ISBB), 2014 IEEE International Symposium on. IEEE; 2014. p. 1–4

[6] Wolpaw JR, Wolpaw EW. Brain-computer interfaces: something new under the sun. In: Wolpaw JR, Winter Wolpaw E, editors. Brain-computer interfaces: principles and practice. Oxford: Oxford University Press; 2012. p. 3–12.

[7] Brain-Computer Interfaces: Applying our Minds to Human-Computer Interaction. Springer Verlag, 2010.

[8] S. Power, "Toward an optical brain-computer interface based on consciously-modulated prefrontal hemodynamic activity." PhD diss., University of Toronto (Canada), 2012.

[9] L.-Y. Chuang, C.-J. Huang, and T.-M. Hung, "The differences in frontal midline theta power between successful and unsuccessful basketball free throws of elite basketball players," International Journal of Psychophysiology, vol. 90, no. 3, pp. 321–328, 2013.

[10] Tan D, Nijholt A. Brain-computer interfaces and human–computer interaction. Brain-computer interfaces. Springer; 2010.

[11] P. Arico, G. Borghini, G. D. Flumeri, A. Colosimo, I. Graziani, J.-P. Imbert, G. Granger, R. Benhacene, M. Terenzi, S. Pozzi, and F. Babiloni, "Reliability over time of EEG-based mental workload evaluation during Air Traffic Management (ATM) tasks," 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015.

[12] P. Arico, G. Borghini, G. D. Flumeri, A. Colosimo, I. Graziani, J.-P. Imbert, G. Granger, R. Benhacene, M. Terenzi, S. Pozzi, and F. Babiloni, "Reliability over time of EEG-based mental workload evaluation during Air Traffic Management (ATM) tasks," 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015

[13] Vecchiato G, Astolfi L, De Vico Fallani F, Salinari S, Cincotti F, Aloise F, Mattia D, Marciani MG, Bianchi L, Soranzo R et al. The study of brain activity during the observation of commercial advertising by using high resolution EEG techniques. In: Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE. IEEE; 2009. p. 57–60.

[14] L. Bai, T. Yu and Y. Li, "A brain computer interface-based explorer", Journal of Neuroscience Methods, vol. 244, pp. 2-7, 2015.

[15] S. Fazli, J. Mehnert, J. Steinbrink, G. Curio, A. Villringer, K. Müller and B. Blankertz, "Enhanced performance by a hybrid NIRS–EEG brain computer interface", NeuroImage, vol. 59, no. 1, pp. 519-529, 2012.

[16] M. Ferrari and V. Quaresima, "A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application," NeuroImage, vol. 63, no. 2, pp. 921–935, 2012.

[17] G. L. Read and I. J. Innis, "Electroencephalography (Eeg)," The International Encyclopedia of Communication Research Methods, pp. 1–18, Jan. 2017.

[18] A. Rowan and E. Tolunsky, A primer of EEG. Philadelphia, PA: Butterworth-Heinemann, 2003.

[19] H.H. Jasper, The ten±twenty electrode system of the International Federation, Electroenceph. clin. Neurophysiol., 10: 371± 375, 1958

[20] V. Chaurasia, V. Mishra, and L. Jain, "Brain-Bot: An Unmanned Ground Vehicle (UGV) Using Raspberry Pi and Brain Computer Interface (BCI) Technology," Communications in Computer and Information Science Advances in Computing and Data Sciences, pp. 252–261, 2017.

[21] S. Niketeghad, 2015. "Towards Closed-loop Deep Brain Stimulation: Behavior Recognition from Human STN" (Doctoral dissertation, University of Denver), 2015.

[22] M. Hosťovecký and B. Babušiak, "Brain activity: beta wave analysis of 2D and 3D serious games using EEG," Journal of Applied Mathematics, Statistics and Informatics, vol. 13, no. 2, 2017.

[23] X. Jia and A. Kohn, "Gamma Rhythms in the Brain", PLoS Biology, vol. 9, no. 4, p. e1001045, 2011.

[24] Sarason, I.G., Pierce, G.R. and Sarason, B.R. eds., 2014. Cognitive interference: Theories, methods, and findings. Routledge.

[25] M. L. Rothschild and Y. J. Hyun, "Predicting Memory for Components of TV Commercials from EEG," Journal of Consumer Research, vol. 16, no. 4, p. 472, 1990.

[26] F. Gobet, "Expert memory: a comparison of four theories," Cognition, vol. 66, no. 2, pp. 115–152, 1998.

[27] A. M. C. Kelly and H. Garavan, "Human Functional Neuroimaging of Brain Changes Associated with Practice," Cerebral Cortex, vol. 15, no. 8, pp. 1089–1102, 2004.

[28] Hoffmann, U., Vesin, J.M. and Ebrahimi, T., 2007. Recent advances in brain-computer interfaces. In IEEE International Workshop on Multimedia Signal Processing (MMSP07) (No. LTS-CONF-2007-063).

[29] K. Petersson, C. Elfgren, and M. Ingvar, "Dynamic changes in the functional anatomy of thehuman brain during recall of abstract designs related topractice," Neuropsychologia, vol. 37, no. 5, pp. 567–587, 1999.

[30] J. Kassubek, K. Schmidtke, H. Kimmig, C. H. Lücking, and M. W. Greenlee, "Changes in cortical activation during mirror reading before and after training: an fMRI study of procedural learning," Cognitive Brain Research, vol. 10, no. 3, pp. 207–217, 2001.

[31] A. Rabbi, K. Ivanca, A. Putnam, A. Musa, C. Thaden, and R. Fazel-Rezai, "Human performance evaluation based on EEG signal analysis: A prospective review," 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009.

[32] D. Gutiérrez and M. A. Ramírez-Moreno, "Assessing a learning process with functional ANOVA estimators of EEG power spectral densities," Cognitive Neurodynamics, vol. 10, no. 2, pp. 175–183, Jan. 2015.

[33] F. Taya, Y. Sun, G. Borghini, P. Arico, F. Babiloni, A. Bezerianos, and N. V. Thakor, "Training-induced changes in information transfer efficiency of the brain network: A functional connectome approach," 2015

[34] Emotiv. Available from: http://www.emotiv.com.

[35] Y. Liu, X. Hou, O. Sourina, D. Konovessis, and G. Krishnan, "Human Factor Study for Maritime Simulator-Based Assessment of Cadets," Volume 3: Structures, Safety and Reliability, 2016.

[36] Y. Liu, X. Hou, O. Sourina, D. Konovessis, and G. Krishnan, "EEG-based human factors evaluation for maritime simulator-aided assessment," Maritime Technology and Engineering III, pp. 859–864, 2016.

[37] S. Perry, S. Bridges, and M. Burrow, "An analysis of expert/novice EEG coherence from a dental haptic simulator," Brain Stimulation, vol. 8, no. 2, p. 328, 2015.

[38] G. Borghini, P. Aricò, I. Graziani, S. Salinari, Y. Sun, F. Taya, A. Bezerianos, N. V. Thakor, and F. Babiloni, "Quantitative Assessment of the Training Improvement in a Motor-Cognitive Task by Using EEG, ECG and EOG Signals," Brain Topography, vol. 29, no. 1, pp. 149–161, 2015.

[39] G. Borghini, P. Aricò, G. D. Flumeri, N. Sciaraffa, A. Colosimo, M.-T. Herrero, A. Bezerianos, N. V. Thakor, and F. Babiloni, "A New Perspective for the Training Assessment: Machine Learning-Based Neurometric for Augmented Users Evaluation," Frontiers in Neuroscience, vol. 11, 2017.

[40] A. D. Antonio, J. Ramirez, and G. Mendez, "An Agent-Based Architecture for Virtual Environments for Training," Developing Future Interactive Systems.

[41] F. D. Rose, E. A. Attree, B. M. Brooks, D. M. Parslow, and P. R. Penn, "Training in virtual environments: transfer to real world tasks and equivalence to real task training," Ergonomics, vol. 43, no. 4, pp. 494–511, 2000.

[42] M. Glezerman, "Yes, there is a female and a male brain: Morphology versus functionality," Proceedings of the National Academy of Sciences, vol. 113, no. 14, Aug. 2016.

[43] J. N. Giedd, A. Raznahan, K. L. Mills, and R. K. Lenroot, "Review: magnetic resonance imaging of male/female differences in human adolescent brain anatomy," Biology of Sex Differences, vol. 3, no. 1, p. 19, 2012.

[44] Ingalhalikar, M., Smith, A., Parker, D., Satterthwaite, T.D., Elliott, M.A., Ruparel, K., Hakonarson, H., Gur, R.E., Gur, R.C. and Verma, R., 2014. Sex differences in the structural connectome of the human brain. Proceedings of the National Academy of Sciences, 111(2), pp.823-828.

[45] J. M. Noyes and D. P. J. Bruneau, "A self-analysis of the NASA-TLX workload measure," Ergonomics, vol. 50, no. 4, pp. 514–519, 2007.

[46] P. Abhang, B. Gawali and S. Mehrotra, Introduction to EEG- and speech-based emotion recognition.

[47] Analyzing neural time series data. Cambridge, Mass. [u.a]: MIT Press, 2015.

[48] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis," Brain Research Reviews, vol. 29, no. 2-3, pp. 169–195, 1999.

[49] Kohavi, R., 1995, August. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Ijcai (Vol. 14, No. 2, pp. 1137-1145).

[50] G. McLachlan, K. Do and C. Ambroise, Analyzing microarray gene expression data. Hoboken, N.J.: Wiley-Interscience, 2005.

[51] Ali, A., Shamsuddin, S.M. and Ralescu, A.L., 2015. Classification with class imbalance problem: a review. Int. J. Advance Soft Compu. Appl, 7(3).

[52] Shuo Wang and Xin Yao, "Multiclass Imbalance Problems: Analysis and Potential Solutions", IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 42, no. 4, pp. 1119-1130, 2012.

[53] S. D. Power, A. Kushki, and T. Chau, "Automatic single-trial discrimination of mental arithmetic, mental singing and the no-control state from prefrontal activity: toward a three-state NIRS-BCI," BMC Research Notes, vol. 5, no. 1, p. 141, 2012.

[54] S. Wang, J. Gwizdka, and W. A. Chaovalitwongse, "Using Wireless EEG Signals to Assess Memory Workload in the n -Back Task," IEEE Transactions on Human-Machine Systems, vol. 46, no. 3, pp. 424–435, 2016.

[55] R. Bruni and G. Bianchi, "Effective Classification Using a Small Training Set Based on Discretization and Statistical Analysis," IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 9, pp. 2349–2361, Jan. 2015.

[56] D. Plass-Oude Bos, B. Reuderink, B. Laar, H. Gürkök, C. Mühl, M. Poel, A. Nijholt, and D. Heylen, "Brain-Computer Interfacing and Games," in Brain-Computer Interfaces, ser. Human-Computer Interaction Series, D. S. Tan and A. Nijholt, Eds., 2010, ch. 10

[57] A. Poyil and C. Hema, "EEG Feature Extraction in Brain-Mobile Phone Interfaces," British Journal of Applied Science & Technology, vol. 8, no. 5, pp. 506–520, Oct. 2015.

[58] Brouwer A-M, van Erp J, Heylen D, Jensen O, Poel M. Effortless passive BCIs for healthy users in Universal Access in Human-Computer Interaction Design Methods. Tools and Interaction Techniques for eInclusion. Springer; 2013.

[59] B. Şen, M. Peker, A. Çavuşoğlu, and F. V. Çelebi, "A Comparative Study on Classification of Sleep Stage Based on EEG Signals Using Feature Selection and Classification Algorithms," Journal of Medical Systems, vol. 38, no. 3, 2014.

[60] K.-Q. Shen, C.-J. Ong, X.-P. Li, Z. Hui, and E. Wilder-Smith, "A Feature Selection Method for Multilevel Mental Fatigue EEG Classification," IEEE Transactions on Biomedical Engineering, vol. 54, no. 7, pp. 1231–1237, 2007.

[61] G. Han and L. Yang, "Study on the Changes of EEG Signal and Driving Behavior Based on the Driving Simulator," Green Intelligent Transportation Systems Lecture Notes in Electrical Engineering, pp. 409–420, Dec. 2017.

[62] Müller-Putz, G., Scherer, R., Brunner, C., Leeb, R. and Pfurtscheller, G., 2008. Better than random: a closer look on BCI results. International Journal of Bioelectromagnetism, 10(EPFL-ARTICLE-164768), pp.52-55.

[63] 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2014. R. Hasler, N. Perroud, H. B. Meziane, F. Herrmann, P. Prada, P. Giannakopoulos, and M.-P. Deiber, "Attention-related EEG markers in adult ADHD," Neuropsychologia, vol. 87, pp. 120–133, 2016.

[64] J. N. Mak, R. H. M. Chan, and S. W. H. Wong, "Evaluation of mental workload in visual-motor task: Spectral analysis of single-channel frontal EEG," IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society, 2013.

[65] Fukushima M, Inoue A, Niwa T. Emotional evaluation of tv-cm using the fractal dimension and the largest lyapunov exponent. In: Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on. IEEE; 2010. p. 1473–76

[66] 7th International IEEE/EMBS Conference on Neural Engineering (NER), 2015. J. F. Hu, Z. D. Mu, and J. H. Yin, "EEG Feature Extraction Based on Wavelet Decomposition," Applied Mechanics and Materials, vol. 496-500, pp. 2023–2026, 2014.

[67] Sorudeykin KA. An educative brain-computer interface. arXiv preprint arXiv:1003.2660; 2010.

[68] G. D. Flumeri, P. Arico, G. Borghini, A. Colosimo, and F. Babiloni, "A new regression-based method for the eye blinks artifacts correction in the EEG signal, without using any EOG channel," 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016.

[69] A. W. V. D. Kooi, I. J. Zaal, F. A. Klijn, H. L. Koek, R. C. Meijer, F. S. Leijten, and A. J. Slooter, "Delirium Detection Using EEG," Chest, vol. 147, no. 1, pp. 94–101, 2015.

[70] A. Hempel, F. L. Giesel, N. M. G. Caraballo, M. Amann, H. Meyer, T. Wüstenberg, M. Essig, and J. Schröder, "Plasticity of Cortical Activation Related to Working Memory During Training," American Journal of Psychiatry, vol. 161, no. 4, pp. 745–747, 2004.

[71] G. Gomez-Herrero, W. Clercq, H. Anwar, O. Kara, K. Egiazarian, S. Huffel, and W. Paesschen, "Automatic Removal of Ocular Artifacts in the EEG without an EOG Reference Channel," Proceedings of the 7th Nordic Signal Processing Symposium - NORSIG 2006, 2006.

[72] H. U. Amin, A. S. Malik, N. Badruddin, and W.-T. Chooi, "Brain Behavior in Learning and Memory Recall Process: A High-Resolution EEG Analysis," IFMBE Proceedings The 15th International Conference on Biomedical Engineering, pp. 683–686, 2014.

[73] G. Vecchiato, G. Borghini, P. Aricò, I. Graziani, A. G. Maglione, P. Cherubino, and F. Babiloni, "Investigation of the effect of EEG-BCI on the simultaneous execution of flight simulation and attentional tasks," Medical & Biological Engineering & Computing, vol. 54, no. 10, pp. 1503–1513, Aug. 2015.

[74] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," Clinical neurophysiology, 2002

[75] G. Gratton, M. G. Coles, and E. Donchin, "A new method for off-line removal of ocular artifact," Electroencephalography and Clinical Neurophysiology, vol. 55, no. 4, pp. 468–484, 1983.

[76] G. Borghini, P. Aricò, G. D. Flumeri, G. Cartocci, A. Colosimo, S. Bonelli, A. Golfetti, J. P. Imbert, G. Granger, R. Benhacene, S. Pozzi, and F. Babiloni, "EEG-Based Cognitive Control Behaviour Assessment: an Ecological study with Professional Air Traffic Controllers," Scientific Reports, vol. 7, no. 1, Mar. 2017.

[77] A. Bertrand, V. Mihajlovic, B. Grundlehner, C. V. Hoof, and M. Moonen, "Motion artifact reduction in EEG recordings using multi-channel contact impedance measurements," 2013 IEEE Biomedical Circuits and Systems Conference (BioCAS), 2013.

[78] B. Berger, S. Omer, T. Minarik, A. Sterr, and P. Sauseng, "Interacting Memory Systems—Does EEG Alpha Activity Respond to Semantic Long-Term Memory Access in a Working Memory Task?," Biology, vol. 4, no. 1, pp. 1–16, 2014.

[79] Y. Wang and S. Makeig, "Predicting Intended Movement Direction Using EEG from Human Posterior Parietal Cortex," Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience Lecture Notes in Computer Science, pp. 437–446, 2009.

[80] H. Ayaz, M. P. Cakir, K. Izzetoglu, A. Curtin, P. A. Shewokis, S. C. Bunce, and B. Onaral, "Monitoring expertise development during simulated UAV piloting tasks using optical brain imaging," 2012 IEEE Aerospace Conference, 2012.

[81] A. Mognon, J. Jovicich, L. Bruzzone, and M. Buiatti, "ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features," Psychophysiology, vol. 48, no. 2, pp. 229–240, Jun. 2011.

[82] H. Luchsinger, Ø. Sandbakk, M. Schubert, G. Ettema, and J. Baumeister, "A Comparison of Frontal Theta Activity During Shooting among Biathletes and Cross-Country Skiers before and after Vigorous Exercise," Plos One, vol. 11, no. 3, 2016.

[83] . N. R. Draper and D. M. Stoneman, "Estimating Missing Values in Unreplicated Two-Level Factorial and Fractional Factorial Designs," Biometrics, vol. 20, no. 3, p. 443, 1964.

[84] C. Babiloni, G. Stella, P. Buffo, F. Vecchio, P. Onorati, C. Muratori, S. Miano, F. Gheller, L. Antonaci, G. Albertini, and P. M. Rossini, "Cortical sources of resting state EEG rhythms are abnormal in dyslexic children," Clinical Neurophysiology, vol. 123, no. 12, pp. 2384–2391, 2012.

[85] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis," Brain Research Reviews, vol. 29, no. 2-3, pp. 169–195, 1999.

[86] P. Petrantonakis and L. Hadjileontiadis, "Emotion Recognition From EEG Using Higher Order Crossings," IEEE Transactions on Information Technology in Biomedicine, vol. 14, no. 2, pp. 186–197, 2010.

[87] A. Varnavas and M. Petrou, "Performance prediction using EEG and trial-invariant characteristic signals," 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2008.

[88] M. Honda, "Dynamic cortical involvement in implicit and explicit motor sequence learning. A PET study," Brain, vol. 121, no. 11, pp. 2159–2173, Jan. 1998.

[89] George, L. and Lécuyer, A., 2010, October. An overview of research on" passive" brain-computer interfaces for implicit human-computer interaction. In International Conference on Applied Bionics and Biomechanics ICABB 2010-Workshop W1" Brain-Computer Interfacing and Virtual Reality".

[90] S. Abdulkader, A. Atia and M. Mostafa, "Brain computer interfacing: Applications and challenges", Egyptian Informatics Journal, vol. 16, no. 2, pp. 213-230, 2015.

[91] B. Hoffman, "Cognitive efficiency: A conceptual and methodological comparison," Learning and Instruction, vol. 22, no. 2, pp. 133–144, 2012.

[92] F. Kirschner, F. Paas, P. A. Kirschner, and J. Janssen, "Differential effects of problem-solving demands on individual and collaborative learning outcomes," Learning and Instruction, vol. 21, no. 4, pp. 587–599, 2011.

[93] E. A. Karuza, L. L. Emberson, and R. N. Aslin, "Combining fMRI and behavioral measures to examine the process of human learning," Neurobiology of Learning and Memory, vol. 109, pp. 193–206, 2014.

[94] G. F. Slenes, G. C. Beltramini, F. O. Lima, L. M. Li, and G. Castellano, "The use of fMRI for the evaluation of the effect of training in motor imagery BCI users," 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER), 2013.

[95] R. A. Poldrack, "Imaging Brain Plasticity: Conceptual and Methodological Issues— A Theoretical Review," NeuroImage, vol. 12, no. 1, pp. 1–13, 2000.

[96] M. J. Khan and K.-S. Hong, "Hybrid EEG–fNIRS-Based Eight-Command Decoding for BCI: Application to Quadcopter Control," Frontiers in Neurorobotics, vol. 11, 2017.

[97] C. Brunner, N. Birbaumer, B. Blankertz, C. Guger, A. Kübler, D. Mattia, J. D. R. Millán, F. Miralles, A. Nijholt, E. Opisso, N. Ramsey, P. Salomon, and G. R. Müller-Putz, "BNCI Horizon 2020: towards a roadmap for the BCI community," Brain-Computer Interfaces, vol. 2, no. 1, pp. 1–10, Feb. 2015.

[98] I. Wiegand, K. Hennig-Fast, B. Kilian, H. J. Müller, T. Töllner, H.-J. Möller, R. R. Engel, and K. Finke, "EEG correlates of visual short-term memory as neuro-cognitive endophenotypes of ADHD," Neuropsychologia, vol. 85, pp. 91–99, 2016.

[99] T. O. Zander, C. Kothe, S. Jatzev, and M. Gaertner, "Enhancing Human-Computer Interaction with Input from Active and Passive Brain-Computer Interfaces," Brain-Computer Interfaces Human-Computer Interaction Series, pp. 181–199, 2010.

[100] S. Abdulkader, A. Atia and M. Mostafa, "Brain computer interfacing: Applications and challenges", Egyptian Informatics Journal, vol. 16, no. 2, pp. 213-230, 2015.

[101] S. Abdulkader, A. Atia and M. Mostafa, "Brain computer interfacing: Applications and challenges", Egyptian Informatics Journal, vol. 16, no. 2, pp. 213-230, 2015.

[102] S. K. Hadjidimitriou and L. J. Hadjileontiadis, "Toward an EEG-Based Recognition of Music Liking Using Time-Frequency Analysis," IEEE Transactions on Biomedical Engineering, vol. 59, no. 12, pp. 3498–3510, 2012.

[103] Imbert, J.P., Granger, G., Benhacene, R., Golfetti, A., Bonelli, S. and Pozzi, S., Skill, Rule and Knowledge-based Behaviors Detection during Realistic ATM Simulations by Means of ATCOs' Brain Activity.

[104] N. Derakshan and M. Eysenck, "Anxiety, Processing Efficiency, and Cognitive Performance", European Psychologist, vol. 14, no. 2, pp. 168-176, 2009.

[105] U. Herwig, T. Kaffenberger, C. Schell, L. Jäncke, and A. B. Brühl, "Neural activity associated with self-reflection," BMC Neuroscience, vol. 13, no. 1, p. 52, 2012.

[106] P. V. D. F. Paiva, L. S. Machado, and A. M. G. Valenca, "A Virtual Environment for Training and Assessment of Surgical Teams," 2013 XV Symposium on Virtual and Augmented Reality, 2013.

[107]   J. Moskaliuk, J. Bertram, and U. Cress, "Training in virtual environments: putting theory into practice," Ergonomics, vol. 56, no. 2, pp. 195–204, 2013.

[108]   R. R. Johnson, B. T. Stone, C. M. Miranda, B. Vila, L. James, S. M. James, R. F. Rubio, and C. Berka, "Identifying psychophysiological indices of expert vs. novice performance in deadly force judgment and decision making," Frontiers in Human Neuroscience, vol. 8, 2014.

[109]   S. K. Hadjidimitriou and L. J. Hadjileontiadis, "EEG-Based Classification of Music Appraisal Responses Using Time-Frequency Analysis and Familiarity Ratings," IEEE Transactions on Affective Computing, vol. 4, no. 2, pp. 161–172, 2013.

[110]   J. Stone, Independent component analysis. Cambridge, Mass.: MIT Press, 2004.

[111]   M. Doppelmayr, T. Finkenzeller, and P. Sauseng, "Frontal midline theta in the pre-shot phase of rifle shooting: Differences between experts and novices," Neuropsychologia, vol. 46, no. 5, pp. 1463–1467, 2008.

[112]   G. Seni and J. F. Elder, "Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions," Synthesis Lectures on Data Mining and Knowledge Discovery, vol. 2, no. 1, pp. 1–126, 2010.

[113]   A. J. Haufler, T. W. Spalding, D. S. Maria, and B. D. Hatfield, "Neuro-cognitive activity during a self-paced visuospatial task: comparative EEG profiles in marksmen and novice shooters," Biological Psychology, vol. 53, no. 2-3, pp. 131–160, 2000.

[114]   G. Bernardi, E. Ricciardi, L. Sani, A. Gaglianese, A. Papasogli, R. Ceccarelli, F. Franzoni, F. Galetta, G. Santoro, R. Goebel, and P. Pietrini, "How Skill Expertise Shapes the Brain Functional Architecture: An fMRI Study of Visuo-Spatial and Motor Processing in Professional Racing-Car and Naïve Drivers," PLoS ONE, vol. 8, no. 10, 2013.

[115]   J. Bhattacharya and H. Petsche, "Drawing on minds canvas: Differences in cortical integration patterns between artists and non-artists," Human Brain Mapping, vol. 26, no. 1, pp. 1–14, 2005.

[116]   E. Noh, G. Herzmann, T. Curran, and V. R. D. Sa, "Using single-trial EEG to predict and analyze subsequent memory," NeuroImage, vol. 84, pp. 712–723, 2014.

[117]   S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, "EEG-Based Emotion Recognition Using Deep Learning Network with Principal Component Based Covariate Shift Adaptation," The Scientific World Journal, vol. 2014, pp. 1–10, 2014.

[118]   J. Desmond and J. Fiez, "Neuroimaging studies of the cerebellum: language, learning and memory", Trends in Cognitive Sciences, vol. 2, no. 9, pp. 355-362, 1998.

[119]   D. Schomer and F. Lopes da Silva, Niedermeyer's electroencephalography.

[120]   T. Nomura and Y. Mitsukura, "EEG-Based Detection of TV Commercials Effects," Procedia Computer Science, vol. 60, pp. 131–140, 2015.

[121]   Royer AS, Doud AJ, Rose ML, He B. Eeg control of a virtual helicopter in 3-dimensional space using intelligent control strategies. Neural Syst Rehabilitation Eng, IEEE Trans 2010;18(6):581-9.

[122]   J. Westwood, Medicine Meets Virtual Reality 20. Amsterdam: IOS Press, 2013.

[123]   C. Guerrero-Mosquera, M. Verleysen, and A. N. Vazquez, "EEG feature selection using mutual information and support vector machine: A comparative analysis," 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, 2010.

[124]   G. Ruthenbeck and K. Reynolds, "Virtual reality for medical training: the state-of-the-art", Journal of Simulation, vol. 9, no. 1, pp. 16-26, 2015.

[125]   A. W. H. House, J. Smith, S. Mackinnon, and B. Veitch, "Interactive simulation for training offshore workers," 2014 Oceans - St. Johns, 2014.

[126]   P. Lightbown, "Discovering Golfs Innermost Truths: A New Approach to Teaching the Game," International Journal of Sports Science & Coaching, vol. 5, no. 2_suppl, pp. 77–87, 2010.

[127]   35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2013. N.-H. Liu, C.-Y. Chiang, and H.-C. Chu, "Recognizing the Degree of Human Attention Using EEG Signals from Mobile Sensors," Sensors, vol. 13, no. 8, pp. 10273–10286, Sep. 2013.

[128]   M.-Y. Cheng, C.-L. Hung, C.-J. Huang, Y.-K. Chang, L.-C. Lo, C. Shen, and T.-M. Hung, "Expert-novice differences in SMR activity during dart throwing," Biological Psychology, vol. 110, pp. 212–218, 2015.

[129]   A. M. Proverbio, N. Crotti, M. Manfredi, R. Adorni, and A. Zani, "Who needs a referee? How incorrect basketball actions are automatically detected by basketball players brain," Scientific Reports, vol. 2, no. 1, 2012.

[130]   T. F. Collura, N. L. Wigton, C. Zalaquett, S. Chatters-Smith, and R. J. Bonnstetter, "The Value of EEG-Based Electromagnetic Tomographic Analysis in Human Performance and Mental Health," *Biofeedback*, vol. 44, no. 2, pp. 58–65, 2016.

[131]   D. Wang, D. Miao, and C. Xie, "Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection," Expert Systems with Applications, 2011.

[132]   M. J. Khan, M. J. Hong, and K.-S. Hong, "Decoding of four movement directions using hybrid NIRS-EEG brain-computer interface," Front. Hum. Neurosci. 8, 244 (2014).

[133]   N. Naseer, M. J. Hong, and K.-S. Hong, "Online binary decision decoding using functional near-infrared spectroscopy for the development of brain-computer interface," Exp. Brain Res. 232(2), 555–564 (2014).

[134]   A. Turnip, K.-S. Hong, and M.-Y. Jeong, "Real-time feature extraction of P300 component using adaptive nonlinear principal component analysis," Biomed. Eng. Online 10(1), 83 (2011)

[135]   Yoshioka M, Inoue T, Ozawa J. Brain signal pattern of engrossed subjects using near infrared spectroscopy (nirs) and its application to tv commercial evaluation. In: Neural Networks (IJCNN), The 2012 International Joint Conference on. IEEE; 2012. p. 1–6.

[136]   Vecchiato G, Babiloni F, Astolfi L, Toppi J, Cherubino P, Dai J, Kong W, Wei D. Enhance of theta eeg spectral activity related to the memorization of commercial advertisings in chinese and italian subjects. In: Biomedical Engineering and Informatics (BMEI), 2011 4th International Conference on, vol. 3. IEEE; 2011. p. 1491–94.

[137]   S. D. Muthukumaraswamy, "High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations," Frontiers in Human Neuroscience, vol. 7, 2013.

[138]   M. Zaepffel, R. Trachel, B. E. Kilavik, and T. Brochier, "Modulations of EEG Beta Power during Planning and Execution of Grasping Movements," PLoS ONE, vol. 8, no. 3, 2013.

[139]   A.-M. Brouwer, M. A. Hogervorst, J. B. F. V. Erp, T. Heffelaar, P. H. Zimmerman, and R. Oostenveld, "Estimating workload using EEG spectral power and ERPs in the n-back task," Journal of Neural Engineering, vol. 9, no. 4, p. 045008, 2012.

[140]   M. Deriche and A. Al-Ani, "A new algorithm for EEG feature selection using mutual information," 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221).

[141]   M. Taghizadeh-Sarabi, K. Niksirat, S. Khanmohammadi, and M. Nazari, "EEG-based analysis of human driving performance in turning left and right using Hopfield neural network," SpringerPlus, vol. 2, no. 1, p. 662, 2013.

[142]   O. Sourina, Y. Liu, and M. K. Nguyen, "Real-time EEG-based emotion recognition for music therapy," Journal on Multimodal User Interfaces, vol. 5, no. 1-2, pp. 27–35, Jan. 2011.

[143]   W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis," Brain Research Reviews, vol. 29, no. 2-3, pp. 169–195, 1999.

[144]   C. Muhl, C. Jeunet, and F. Lotte, "EEG-based workload estimation across affective contexts," Frontiers in Neuroscience, vol. 8, Dec. 2014.

[145]   J. J. Larocque, J. A. Lewis-Peacock, A. T. Drysdale, K. Oberauer, and B. R. Postle, "Decoding Attended Information in Short-term Memory: An EEG Study," Journal of Cognitive Neuroscience, vol. 25, no. 1, pp. 127–142, 2013.

[146]   F. Al-Shargie, M. Kiguchi, N. Badruddin, S. C. Dass, A. F. M. Hani, and T. B. Tang, "Mental stress assessment using simultaneous measurement of EEG and fNIRS," Biomedical Optics Express, vol. 7, no. 10, p. 3882, Jun. 2016.

[147]   "Author-Initiated Retraction: Anderson et al, Induced Alpha Rhythms Track the Content and Quality of Visual Working Memory Representations with High Temporal Precision," Journal of Neuroscience, vol. 35, no. 6, pp. 2838–2838, Nov. 2015.

Appendices

Appendix A : Classification accuracy table for adjacent trials with balanced classes

	1 vs 2	2 vs 3	3 vs 4	4 vs 5	5 vs 6	6 vs 7	7 vs 8	8 vs 9	9 vs 10
Part 1	87.7	82.8	82.9	66.6	67.5	68.1	77.1	77.4	74.3
Part 2	94.7	97.7	73.0	70.1	85.9	81.1	87.9	90.5	78.3
Part 3	76.2	74.0	82.3	69.6	78.6	78.9	68.3	64.7	84.3
Part 4	89.3	87.6	84.3	99.1	94.4	76.9	81.7	91.2	81.4
Part 5	89.8	92.4	80.5	88.2	90.8	87.0	78.4	73.7	74.9
Part 6	67.6	68.8	74.2	77.6	62.1	59.8	68.4	69.4	68.6
Part 7	76.3	72.8	78.7	83.8	73.9	76.2	77.1	81.5	67.8
Part 8	72.3	73.9	65.1	64.8	64.2	66.9	68.2	69.7	70.5
Part 9	74.9	70.1	80.9	76.2	72.8	77.1	67.2	64.6	61.1
Part 10	74.8	67.2	71.9	68.7	74.4	68.4	65.5	71.1	69.7
Mean	80.4	78.7	77.4	76.5	76.5	74.0	74.0	75.4	73.1

Appendix B : Number of epochs used in the classification per participant per trial

	Participant 1		
Trial	Time taken to complete the trial in seconds	Number of data points after noise removal	Number of epochs used in classification
Participant 1			
Trial 1	163	51889	51
Trial 2	218	64569	64
Trial 3	217	84075	84
Trial 4	160	72342	72
Trial 5	223	108234	108
Trial 6	233	104270	104
Trial 7	217	104290	104
Trial 8	211	98444	98
Trial 9	182	89548	89
Trial 10	211	97933	97
Participant 2			
Trial 1	334	160009	160
Trial 2	313	140590	140
Trial 3	306	134517	134
Trial 4	315	124808	124
Trial 5	304	109344	109
Trial 6	312	113398	113
Trial 7	305	81974	81
Trial 8	305	106112	106
Trial 9	286	88957	88
Trial 10	290	102761	102
Participant 3			
Trial 1	207	31379	31
Trial 2	235	98632	98
Trial 3	197	83551	83
Trial 4	214	61014	61
Trial 5	196	81946	81
Trial 6	206	65826	65
Trial 7	205	77997	77

Trial 8	191	73760	73
Trial 9	209	74420	74
Trial 10	180	60792	60
Participant 4			
Trial 1	182	85219	85
Trial 2	188	80908	80
Trial 3	163	68779	68
Trial 4	209	91443	91
Trial 5	199	72468	72
Trial 6	160	70598	70
Trial 7	208	82770	82
Trial 8	178	58542	58
Trial 9	200	53263	53
Trial 10	215	56986	56
Participant 5			
Trial 1	174	41170	41
Trial 2	196	66324	66
Trial 3	158	57542	57
Trial 4	174	50479	50
Trial 5	162	48306	48
Trial 6	164	61960	61
Trial 7	162	45063	45
Trial 8	167	59988	59
Trial 9	172	49800	49
Trial 10	180	46355	46
Participant 6			
Trial 1	281	128384	128
Trial 2	300	142366	142
Trial 3	314	131112	131
Trial 4	379	161303	161
Trial 5	262	126640	126
Trial 6	266	120273	120
Trial 7	302	142318	142
Trial 8	316	141895	141
Trial 9	297	140657	140
Trial 10	308	131955	131
Participant 7			
Trial 1	261	115347	115
Trial 2	262	116954	116
Trial 3	233	100123	100

Trial 4	237	97904	97
Trial 5	205	90030	90
Trial 6	213	97431	97
Trial 7	206	87110	87
Trial 8	216	99844	99
Trial 9	190	87173	87
Trial 10	209	96593	96
Participant 8			
Trial 1	216	85855	85
Trial 2	204	86622	86
Trial 3	230	111570	111
Trial 4	196	88507	88
Trial 5	267	131601	131
Trial 6	279	136883	136
Trial 7	251	120703	120
Trial 8	242	119060	119
Trial 9	274	120723	120
Trial 10	247	121517	121
Participant 9			
Trial 1	227	104091	104
Trial 2	248	113710	113
Trial 3	266	128456	128
Trial 4	276	133480	133
Trial 5	289	141018	141
Trial 6	292	117235	117
Trial 7	285	113919	113
Trial 8	290	99911	99
Trial 9	290	136475	136
Trial 10	270	126870	126
Participant 10			
Trial 1	246	114225	114
Trial 2	260	120604	120
Trial 3	283	117011	117
Trial 4	262	122907	122
Trial 5	240	111680	111
Trial 6	200	97083	97
Trial 7	206	99533	99
Trial 8	211	100856	100
Trial 9	301	148052	148
Trial 10	173	83015	83