EEG-based classification of visual and auditory monitoring tasks

By

Mohammad Bagheri

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

St. John’s, Newfoundland

ABSTRACT

Using EEG signals for mental workload detection has received particular attention in passive BCI research aimed at increasing safety and performance in high-risk and safety-critical occupations, like pilots and air traffic controllers. Along with detecting the level of mental workload, it has been suggested that being able to automatically detect the type of mental workload (e.g., auditory, visual, motor, cognitive) would also be useful. In this work, a novel experimental protocol was developed in which subjects performed a task involving one of two different types of mental workload (specifically, auditory and visual), each under two different levels of task demand (easy and difficult). The tasks were designed to be nearly identical in terms of visual and auditory stimuli, and differed only in the type of stimuli the subject was monitoring/attending to. EEG power spectral features were extracted and used to train linear and non-linear classifiers. Preliminary results on six subjects suggested that the auditory and visual tasks could be distinguished from one another, and individually from a baseline condition (which also contained nearly identical stimuli that the subject did not need to attend to at all), with accuracy significantly exceeding chance. This was true when classification was done within a workload level, and when data from the two workload levels were combined. Preliminary results also showed that tasks with easy and difficult trials could be distinguished from one another, each within a sensory domain (auditory and visual) as well as with both domains combined. Though further investigation is required, these preliminary results are promising, and suggest the feasibility of a passive BCI for detecting both type and level of mental workload.

ACKNOWLEDGEMENTS

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

Thanks a lot to my fellow graduate students: Mahsa Bagheri, Faghihe Massaeli, and Roy Brushett for their help and support. Thanks to all of my friends and seniors for being so supportive and helpful.

TABLE OF CONTENTS

List of figures v

List of tables vi

List of Abbreviations vii

Chapter 1: Introduction 1

1.1. Problem statement 1

1.2. Research Objectives 2

1.3. Thesis Organization 3

Chapter 2: Literature review 5

2.1. Brain Computer Interfaces (BCI) 5

2.2. Passive BCI for mental workload detection 6

2.3. BCI algorithm 8

2.4. Beyond detection of mental workload level: considering Multiple Research Theory 21

2.5. Relevant studies in the literature 23

Chapter 3: Methods 25

3.1. Study Participants 25

3.2. Instrumentation 26

3.3. Study Design ... 28

3.4. Data Analysis ... 36

Chapter 4: Results ... 45

4.1. Validation of induction of workload levels ... 45

4.2. EEG-based state classification ... 46

Chapter 5: Discussion ... 56

Chapter 6: Conclusions ... 60

6.1. Study conclusions ... 60

6.2. Study limitations ... 60

6.3. Future work ... 61

References ... 62

Appendices ... 74

List of figures

Figure 1 Mental workload detection algorithm 9

Figure 2 Standard 64Ch electrode positions used in this work based on 10-10 Standardization 13

Figure 3 EEG frequency bands 15

Figure 4 Separating two classes (black and white) with three different linear decision boundaries 19

Figure 5 a) Triggerbox b) ActiChamp Amplifier 27

Figure 6 Electrodes positioning and cap used in this work 27

Figure 7 a) A sample of English alphabet stimulus presented in this work b) A sample of Greek alphabet stimulus presented in this work 30

Figure 8 Timeline of experimental session and an example task block (Eyes O-C = block with one eyes-open and one eyes-closed baseline trial) 33

Figure 9 Rating Scale Mental Effort © 2020 IEEE 35

Figure 10 Timeline of a single trial 35

Figure 11 ICA Components of brain and non-brain activities 39

Figure 12 EEG signals a) before preprocessing, b) after EMG removal but before blink removal, and c) after all pre-processing steps 40

Figure 13 Classification accuracies (averaged across all classification problems considered) for different feature set sizes. 47

List of tables

Table 1 The types of trials completed during the experimental session © 2020 IEEE 32

Table 2 Mean task performance ratings for each task type and workload level 46

Table 3 Mean RSME ratings for each task type and workload level 46

Table 4 Across-subject average classification accuracies for the Auditory vs. Visual classification problems.. 48

Table 5 Per-participant accuracies for the Auditory vs. Visual classification problems. All accuracies were obtained with the regularized LDA (rLDA) classifier and a 25-dimensional feature set selected from the feature pool that included 56 frequency bands. 50

Table 6 Across-subject average classification accuracies for the Easy vs. Difficult classification problems.. 51

Table 7 Per-participant accuracies for the Easy vs. Difficult classification problems. All accuracies were obtained with the regularized LDA (rLDA) classifier and a 25-dimensional feature set selected from the feature pool that included 56 frequency bands. 52

Table 8 Across-subject average classification accuracies for all individual task conditions vs. corresponding baseline (with stimuli) 53

Table 9 Per-participant accuracies for all individual task condition vs. baseline classification problems. All accuracies were obtained with rLDA classifier and a 25-dimensional feature set selected from the feature pool that included 56 frequency bands. 54

List of Abbreviations

AD: Auditory Difficult

AE: Auditory Easy

ANOVA: Analysis of Variance

B: Baseline (without stimuli)

BCI: Brain Computer Interface

BD: Baseline Difficult

BE: Baseline Easy

COVID-19: Coronavirus Disease of 2019

ECG: Electrocardiogram

ECoG: Electrocorticography

EEG: Electroencephalography

EMG: Electromyography

EOG: Electrooculography

fMRI: functional Magnetic Resonance Imaging

FN: False Negative

FP: False Positive

IAF: Individual alpha frequency

ICA: Independent Component Analysis

KNN: K-Nearest Neighbors

LDA: Linear Discriminant Analysis

MATLAB: Matrix Laboratory

MEG: Magnetoencephalography

MI: Mutual Information

mRMR: minimum Redundancy Maximum Relevance

MRT: Multiple resource theory

NASA-TLX: NASA-Task Load Index

NIRS: Near-Infrared Spectroscopy

PBCI: Passive Brain Computer Interface

rLDA: Regularized Linear Discriminant Analysis

RSME: Rating Scale Mental Effort

SVM: Support Vector Machine

TN: True Negative

TP: True Positive

VD: Visual Difficult

VE: Visual Easy

Chapter 1: Introduction

1.1. Problem statement

A passive brain-computer interface (BCI) is a system that improves human-computer interaction by providing implicit information on a subject's mental state and adapting the environment accordingly [1]. One potential application of a passive BCI is the monitoring of mental workload [2, 3, 4, 5], especially for safety-critical occupations like pilots, air-traffic controllers, and other industrial operators. The goal is to estimate the cognitive strain on the subject from EEG signal variables so that appropriate adaptation strategies can be engaged to reduce the potential for error during periods of extreme demand or overload. Such a technology can have significant industrial and economic impact by preventing accidents related to operator error, and their associated human, economic, and environmental losses.

While automatic detection of mental workload level based on neural signals has been the focus of many studies so far [6, 7, 8, 9], very few studies have yet explored the detection of the type (e.g., cognitive, visual, auditory, motor) of mental workload [10, 11]. Regardless of the application, users of passive BCI systems in real-life scenarios will likely perform activities involving multiple, disparate tasks engaging different sensory domains. BCI systems would be capable of invoking more appropriate adaptation schemes if they had knowledge of both the workload type and level of the user. For example, a system could provide new information aurally when the user is already engaged in a task involving visual processing. This thesis research represents a preliminary step toward filling this gap in the passive BCI and mental workload detection research.

1.2. Research Objectives

The purpose of this research was to work towards the development of a passive BCI for mental workload detection that can automatically identify both the type and level of workload an individual is experiencing at a given time. As a first step toward this long-term objective, this work focussed on differentiating auditory and visual workload. Visual and auditory modalities were selected (as opposed to cognitive or motor) since they would be relevant to a wide range of realistic scenarios, and also because for the purposes of the experiment it would be easier to design tasks that were similar in every way except the sensory modality. The specific research objectives of this thesis were to:

1) Develop machine learning algorithms to automatically identify when an individual is experiencing workload primarily from the visual domain vs. the auditory domain at a level significantly greater than chance using EEG signals;
2) Determine the effect that varying workload levels within each sensory domain has on the accuracy with which the type of sensory processing (visual or auditory) can be identified using EEG;

3) Develop machine learning algorithms to automatically identify distinct levels of mental workload within the visual and auditory domains at a level significantly greater than chance using EEG signals.

1.3. Thesis Organization

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

Chapter 2 presents the literature review of topics relevant to the thesis research. Specifically, passive brain-computer interfaces and mental workload detection are introduced, and various signal acquisition, signal processing, and classification methods used in BCIs are discussed. Multiple resource theory (MRT) of mental workload is described to help motivate this work. Furthermore, relevant existing studies from the literature are discussed.

Chapter 3 presents a detailed description of how the experiment was done and how the data were collected from 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.

Please note that some portions of this thesis were initially published in [12]:

Mohammad Bagheri, Sarah Power, "EEG-based classification of visual and auditory monitoring tasks," *IEEE International Conference on Systems, Man, and Cybernetics (SMC),* pp. 4032-4037, 2020. © 2021 IEEE

Chapter 2: Literature review

2.1. Brain Computer Interfaces (BCI)

Brain–Computer Interface (BCI) systems establish a direct communication channel from the brain to an output device. Mental activity leads to changes in neurophysiological signals. Brain–computer interfaces measure these signals and transform them into a control signal to enable users to control a variety of applications like video games [13], wheelchairs [14], etc. Several studies have even shown the possibility of using electrical brain activity recorded from sensors placed within the brain to directly control the movement of robots or prosthetic devices [15, 16].

BCI systems that require users to intentionally generate distinct and predictable patterns of neurophysiological signals and thus provide explicit commands to control an external device are called "active" BCIs [17, 18]. The examples above would fall into this category. On the other hand, a "passive" BCI is a BCI which derives its outputs from spontaneous brain activity arising without the purpose of voluntary control. This activity reflects the user's mental (e.g., cognitive or emotional) state, and is used to enrich a human-computer interaction with this implicit information. Passive BCIs normally monitor longer periods of brain activity for the detection of a cognitive state change or emotional arousal [17, 19]. An example of this is a system that monitors a driver's neural dynamics in real-time and alarms him/her in the case of drowsiness detection [20].

Passive BCIs have found application in various fields of research, including medicine, neuroergonomics and smart environments, neuromarketing and advertisement, education, games and entertainment, and security and authentication [21, 22]. One of the most widely studied applications in passive BCIs is mental workload detection.

2.2. Passive BCI for mental workload detection

Mental workload consist of the combination of mental effort, information processing, and emotion in response to task demand. Mental workload can be determined by considering the task and the operator. The task-oriented approach considers just the task characteristics and the condition of task performance in estimating workload. The human-oriented approach, on the other hand, evaluates workload through the effect of the task performance on the person. Psychologists are inclined toward the latter approach, viewing workload as the result of the interaction between work demands and human capacity [23, 24].

A user with high mental workload level is more susceptible to making a mistake, thus it can increase the risk of error in performance in any environment. Using passive BCIs that continuously monitor and detect mental workload level and adjust human-computer interaction/environment accordingly would be useful. For instance, such a BCI could activate autopilot to take over more functionality of navigation in an aircraft when periods of high mental workload are detected.

Traditionally, the most common technique for measuring task-induced mental workload has been subjective rating via specifically designed questionnaires such as the NASA-Task Load Index (NASA-TLX) [25], and Rating Scale Mental Effort (RSME) [26]. When the information about mental workload is needed continuously and in real-time as in a BCI, however, this approach is not viable. The individual would be required to provide self-reports at regular intervals, which would continually interrupt them from the task at hand. This would be frustrating, and the disruption in their attention to the task would likely in itself lead to a degradation in performance. While this is sometimes done in research studies [27, 28, 6], it would be untenable in real-world scenarios. Having an objective measure of mental workload based on physiological signals would eliminate these issues, and provide a convenient and unobtrusive method for monitoring mental workload in real-time.

Physiological correlates of workload have been well-studied in the literature. Some of these measures include heart rate, blood pressure, heart rate variability, eye blink frequency, saccades, electromyography (EMG) and electroencephalography (EEG) [29, 21]. While there is no firm consensus on the best physiological indicator of workload, some studies have shown EEG to be more promising compared to other indicators [30, 31, 32]. Mental workload detection via EEG has been studied in relation to many different applications such as aircraft piloting, air traffic control, and car driving applications [33, 34, 35, 36, 37].

2.3. BCI algorithm

Typically, a BCI system, whether active or passive, consists of six basic components (see Figure 1). These include 1) signal acquisition, 2) signal preprocessing, 3) feature calculation, 4) feature selection, 5) classification, and 6) feedback to the user. The signal acquisition component is responsible for recording the neural signals and sending them to the pre-processing component for signal enhancement and noise reduction. The feature calculation components computes a pool of signal characteristics which will potentially allow for discrimination between the mental states of interest. The feature selection component reduces the dimensionality of this overall feature pool, automatically selecting only the most discriminative features for the scenario at hand. The classification component is responsible for predicting the mental state of the user based on the selected signal features. An appropriate command is then sent to the connected system based on the predicted mental state. For instance, if the system detects a low mental workload state it might send a command to the autopilot navigation system to reduce the degree of automation and allow the user control over more tasks. The resulting change in the device/system comprises the feedback to the user. The following sections will provide a more detailed description of each of these BCI algorithm components.

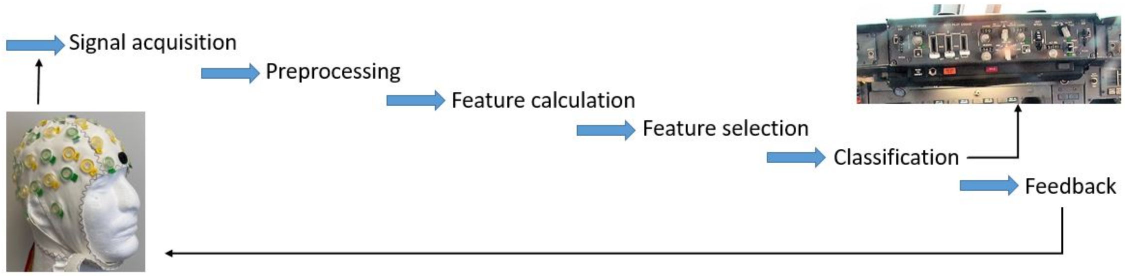


Figure 1 Mental workload detection algorithm

2.3.1. Signal acquisition

EEG signal acquisition can be done in two ways: using invasive or non-invasive methods. In invasive technologies, electrodes are surgically implanted either into the user's brain tissue or on the surface of the brain, while in non-invasive technologies, the brain activity is measured using external sensors placed in contact with the scalp.

Invasive signal acquisition methods provide vastly superior signal quality to non-invasive methods, with relatively high temporal and spatial resolution, as well as signal-to-noise ratios. However, due to the risks associated with invasive technologies, use of these techniques are limited to applications where the benefit might outweigh these risks – for example, for the restoration of communication to individuals with late stage amyotrophic lateral Sclerosis (ALS), or prosthetic control for patients with quadriplegia. Invasive sensor technologies are not currently appropriate for applications aimed at healthy individuals, including for mental workload detection applications [38, 39, 40].

There are several non-invasive functional brain imaging technologies that have been explored for use in BCIs, such as functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), near-infrared spectroscopy, and electroencephalography (EEG). Signals acquired via MEG are extremely small, several orders of magnitude smaller than other signals in a typical environment that can obscure the signal [41] [38]. fMRI is very expensive and the imaging procedure is very restrictive, making this technology unsuitable for BCI applications, which must be able to be used in real-world settings during everyday tasks [42] [38]. NIRS and EEG have emerged as the most promising signal acquisition technologies for BCI. While NIRS has better spatial resolution as compared to EEG, the much superior temporal resolution of EEG has made it more popular for BCI applications, which must function in as close to real-time as possible [38].

2.3.1.1. Electroencephalography (EEG)

Electroencephalography (EEG) measures the electrical activity of the brain using electrodes placed on the scalp. The electrodes are typically attached to the scalp using a tight, flexible cap. The electrodes detect small electrical charges which result from the activity of neurons, mostly in the cortical tissue of the brain. EEG has unique usability advantages over other types of brain signal recording. In particular, EEG is easy to use, portable and relatively inexpensive. EEG recording also provides high temporal resolution, though its spatial resolution is inferior compared to other methods [43].

2.3.2. EEG recording techniques

Electroencephalographic measurements employ a recording system consisting of:

- Electrodes with conductive media: Read the signal from the surface of the scalp. A conductive media (typically, an electrolyte gel) is used to reduce the electrical impedance between the electrode and the scalp, increasing signal quality.
- Amplifier with filter: Enhances the small EEG signals from the microvolt range into the range where they can be digitized accurately.
- A/D converter: Converts the analog electrical signals to digital form.
- Recording device (e.g., computer): Stores and displays obtained digital signal data.

Proper functioning of the EEG electrodes are critical for acquiring sufficiently high quality data for effective interpretation. Many types of electrodes exist, with different characteristics. For instance, active electrodes have a pre-amplification module immediately after the conductive material between the skin and the electrode. These electrodes amplify EEG signals before additional noise can be added as the signal travels between the electrode and the main amplifier. Passive electrodes do not have this pre-amplification module, and thus there is risk that the signal will be contaminated with noise before reaching the amplifier. Active electrodes are recommended, particularly when the individual being monitored is in motion, as the movements and muscle activity can generate artifacts [44].

2.3.3. Standardization of electrode placement

The International Federation of Clinical Neurophysiology adopted standardization for the designation and physical placement of EEG electrodes on the scalp, called the 10-20 System of Electrode Placement [45]. The head is divided into proportional distances from prominent skull landmarks to provide adequate coverage of all regions of the brain. Label 10-20 designates proportional distance between adjacent electrodes which are either 10% or 20% of the total front–back or right–left distance of the skull [46]. For higher resolution systems (i.e., systems with more electrodes), modified versions of the 10-20 system are used where additional electrodes are placed at intermediate sites halfway between those of the existing 10–20 system. In this work, a standard 64-channel electrode placement, following the “10-10 system” (see Figure 2) has been used.

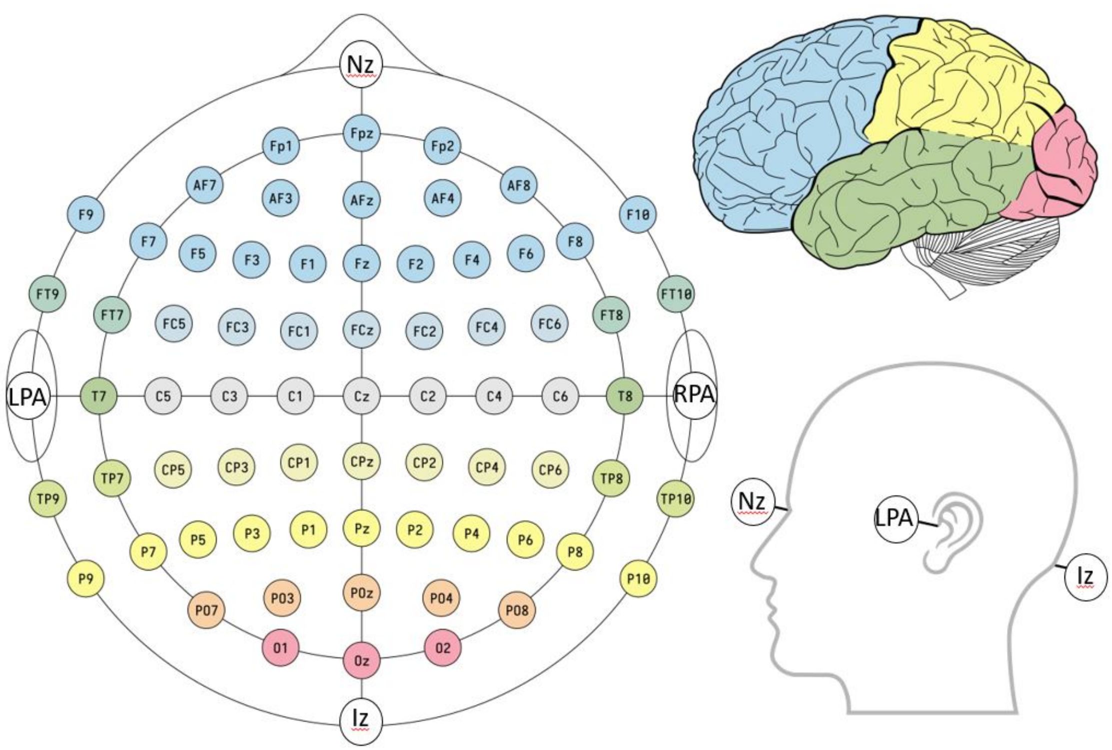


Figure 2 Standard 64Ch electrode positions used in this work based on 10-10 Standardization

2.3.4. Signal pre-processing

EEG data tends to contain a lot of noise, or artefacts, which can obscure weaker EEG signals. Artefacts in the recorded EEG may be either subject-related or technical/environmental. Subject-related artefacts are electrophysiological signals from non-neural sources that can contaminate the EEG. Technical or environmental artefacts arise through the process of recording the EEG or come from the surrounding environment. The most common EEG artefact sources can be classified in following way:

Subject-related:

- EMG (electrical signals resulting from muscle activation)
- ECG (electrical signals related to cardiac activity)
- Eye movements (blinking and saccades, i.e., horizontal and vertical movements of the eyes)
- Sweating (result in impedance changes between the scalp and electrode)

Technical:

- 50/60 Hz line noise (from power supply)
- Motion artefacts from displacement of electrodes (e.g., due to subject movement)
- Impedance fluctuation (e.g., drying up of conductive gel)
- Electrical interference from nearby equipment in the environment

It is important to mitigate and remove these artefacts as much as possible to sufficiently increase the signal-to-noise ratio of the EEG, and increase the chances of successfully detecting the desired underlying mental state encoded therein.

2.3.5. Feature calculation

Raw EEG signals can be transformed into more informative data by feature calculation. There are many features which can be extracted from EEG signals such as frequency band power features, time point features, covariance, coherence, and connectivity features [47]. Among these features, frequency band power features are most commonly used in BCI studies [47]. The power of EEG signals for a certain frequency band in a specific channel, averaged over a specific time frame, is represented as band power features.

Combination of various types of features can also be used to higher classification accuracies compared to using a single feature type although it increases dimensionality of the feature set.

2.3.5.1. EEG frequency bands

In passive BCI research, which aims to detect cognitive or emotional states, EEG activity reflecting underlying neural oscillations, as opposed to transient activity (e.g., event-related potentials), is of primary interest. Neural oscillations present in the EEG have been divided into specific frequency bands, since activity within these specific bands have been linked to underlying cognitive states. These frequency bands are commonly referred to as the delta, theta, alpha, beta, and gamma bands. They are measured in cycles per second or hertz (Hz) [48, 49]

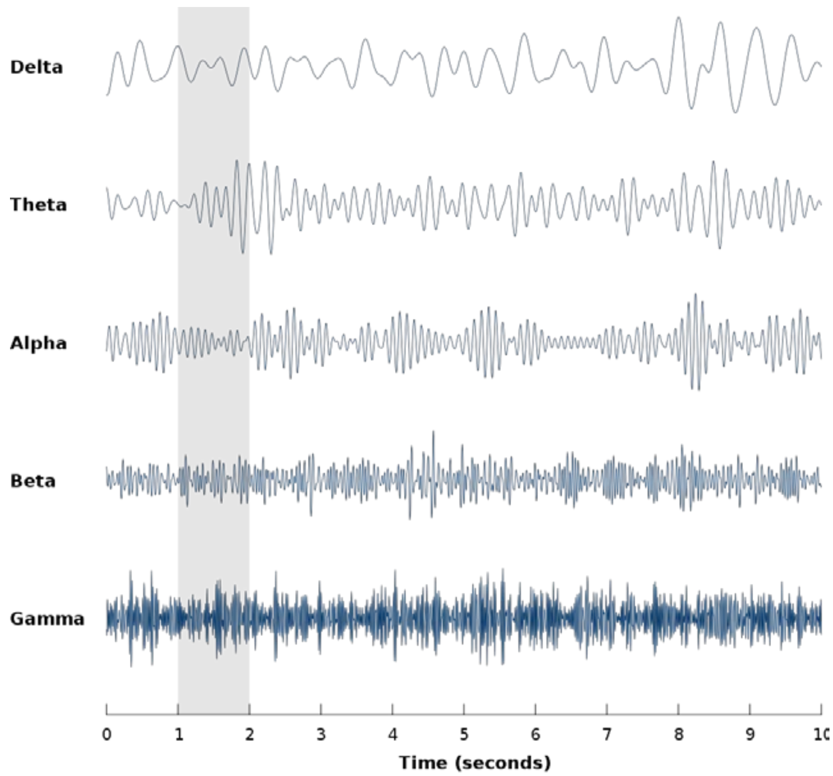


Figure 3 EEG frequency bands

Delta waves (1-3 Hz) are the slowest waves with highest amplitude. These waves are generated when a person is asleep. Delta waves are also known as slow-wave sleep which aid in characterizing the depth of sleep [47].

Theta waves (4-7 Hz) reflect a very relaxed state, representing the zone between waking and sleep [47].

Alpha waves (8-12 Hz) are slower and larger and associated with a state of relaxation. Alpha brainwaves appear upon eye closure [47].

Beta waves (13 – 38 Hz) are small, faster brainwaves associated with a state of mental, intellectual activity and outwardly focused concentration [47].

Gamma brainwaves (39 – 42 Hz) are the fastest and relatively smaller. Gamma rhythms modulate perception and consciousness [47]. Figure 3 shows a short sample of each of EEG waves.

2.3.6. Feature selection

Feature selection is the process of eliminating trivial and redundant features to obtain a smaller subset of informative features that are sufficient for reliable classification. Several approaches have been proposed in the literature to address this step. One effective and commonly used approach, which is used in this work, is the minimum redundancy maximum relevance (mRMR) algorithm.

2.3.6.1. Minimum Redundancy Maximum Relevance (mRMR)

The minimum redundancy maximum relevance approach is based on recognizing that integrating variables that are selected based on how effective they are when considered individually does not necessarily lead to good prediction performance. It aims to reduce the redundancies among the selected variables to a minimum for creating the selected subset of variables [50, 51].

Most of the feature selection algorithms only consider the relationship between the features and the classification categories but ignore the mutual information among features. Instead, the mRMR feature selection algorithm considers not only the amount of information provided by these features for discriminating the states to be classified but also the influence of interaction among features on classification. “Relevance” is the distributional similarity between a continuous feature vector and a target vector, whereas “redundancy” measures the similarity between the distribution of attributes and the distribution of labels.

The redundancy measure utilizes the quantity of mutual information (MI) between two features. If the value of MI is small, it means that there is relatively little information duplication between the features; i.e., there are no significant redundancies between the features. The relevance measure utilizes the value of MI between the feature and the states to be classified. If the value of MI is large, it indicates that there is a strong correlation between the feature and the target states. Therefore, the goal is to minimize the redundancy criterion while maximizing the relevance criterion.

The mRMR algorithm works iteratively. At each iteration, it identifies the best feature and adds it to the feature set by calculating the score for each remaining feature with the already selected features, and adding the best one each time.

In practice, at each iteration i, a score is computed for each feature to be evaluated (f):

$$Score(f) = \frac{\text{relevance } (f \mid target)}{\text{redundancy } (f \mid features\ selected\ until\ i - 1)}$$

The best feature at iteration i is the one having the highest score.

2.3.7. Classification

The most popular classifiers for BCI application are linear classifiers [47]. Linear classifiers use linear decision boundaries to distinguish classes. Figure 4 shows three different linear decision boundaries used for separating two classes of black and white. In this category, linear discriminant analysis (LDA), and support vector machines (SVMs) (with a linear kernel) are commonly used in EEG-based BCIs. Non-linear classifiers, on the other hand, are classifiers with a non-linear decision boundary. Various non-linear classifiers have also been explored for use in BCIs. One common example is the k-nearest neighbor classifier. Since these specific classifiers are commonly used in EEG-based BCI studies [48], they were used in this work and will be described in more detail below.

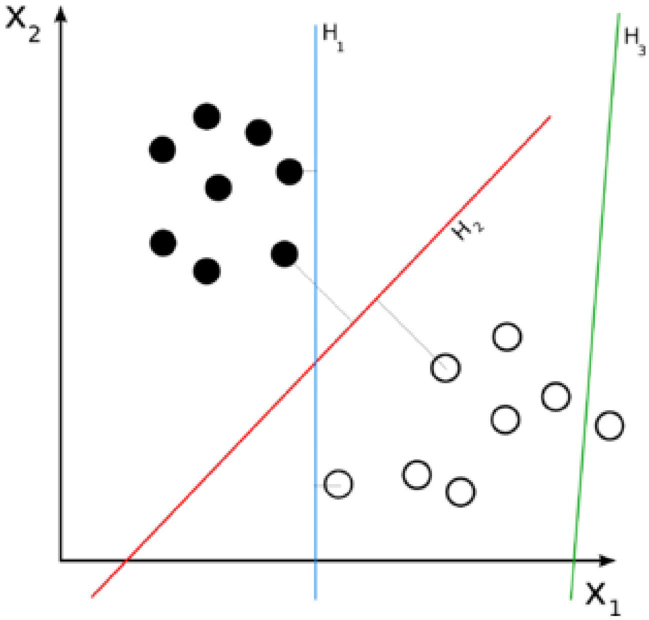


Figure 4 Separating two classes (black and white) with three different linear decision boundaries

2.3.7.1. Linear discriminant analysis (LDA)

LDA uses hyperplanes to separate the data belonging to different classes. LDA assumes a Gaussian distribution of the data and common covariance matrix for different classes [52]. The separating hyperplane is obtained by seeking the projection that maximizes the distance between the two class means and minimizes the interclass variance [53]. This technique has a very low computational requirement which makes it suitable for BCI systems. Moreover, this classifier is simple to use and generally provides good results.

LDA has been used in a variety of passive BCI systems achieving very high accuracy levels [20, 54, 55]. In another study, with the aim of classification of spontaneous EEG during five mental tasks, it has been shown that LDA has comparable results to support vector machines and neural networks [56].

2.3.7.2. Regularized LDA

Regularized Linear Discreminant Analysis (rLDA) was proposed by Friedman [57] to reduce the dispersion of eigenvalues of sample data. The regularization technique is based on replacing within-group sample covariance of data by a weighted average of the whole sample covariance using a shrinking intensity parameter. This parameter can be determined based on the performance of the model on validation data, or by cross-validation. Since rLDA alleviates the degradation of classification accuracy, it is generally known to be appropriate for high dimensional datasets [58]. This approach has been used in several studies with high dimensional and non-stationary problems achieving very high accuracy levels [59, 60].

2.3.7.3. Support Vector Machine

SVM is one of the most commonly used supervised learning algorithm that uses a kernel function to transform input data into higher dimensional space. This classifier is mostly used for binary classification problems in machine learning because of its ability to manage large datasets.

SVM uses a hyperplane to separate target classes [52]. In SVMs, the discriminant hyperplane is the one that maximizes the margin, i.e., the distance from the nearest training points. Maximizing margins enables the SVM to generalize better [52, 61]. SVM has a regularization parameter which makes this algorithm more robust to outliers and noise on the training set; i.e., SVM has good generalization properties and is insensitive to overtraining and to the curse-of-dimensionality [52, 62], which is a challenge in BCI problems, as BCI datasets have low signal to noise ratio. SVM has been a successful classifier in various BCI studies [63, 61, 64, 65].

2.3.7.4. K-nearest neighbours (KNN)

KNN classifier is a discriminative nonlinear classifier which identifies a test sample's class according the majority class of the k (where k is a positive integer) nearest training samples. These k nearest neighbor samples are typically determined using the distance between the feature vectors of each training sample and the test sample [66]. With a sufficiently high value of k and enough training samples, KNN can approximate any function which enables it to produce nonlinear decision boundaries.

2.4. Beyond detection of mental workload level: considering Multiple Research Theory

There is a growing body of literature investigating the feasibility of automatic mental workload detection via neural signals. Many studies have demonstrated the ability to accurately classify patterns of brain activity associated with differing levels of task difficulty in various laboratory tasks [67, 68] as well as in more realistic task scenarios like flight simulation and during driving [4, 5]. Electroencephalography (EEG) is the favoured modality to obtain the neural signals due to its practicality, superior temporal resolution, and relatively low cost [69, 70].

The focus of mental workload studies to date has been almost exclusively on predicting the *level* of mental workload. It has been suggested, however, that knowledge of the *type* of mental workload would also be very valuable [71, 11].

According to the Multiple Resource Theory (MRT) [72, 73, 74, 75, 76], each person has a limited number of resources available for mental activities that are sometimes distinct from one another (for example, it is possible to do multiple attentional tasks at once without them interfering with one another). These resources are regarded as a source of energy used for a wide range of mental functions, from sensation to perception. These resources are allocated across different tasks, modalities, and processing. This theory explains how challenging single-tasks can cause processing problems and how dual-task performance is more likely to be affected by performing similar tasks than dissimilar tasks.

Multiple resource theory [72, 73, 74, 75, 76] suggests that cognitive resources are divided into separate dimensions, which can be thought of as sensory domains: cognitive, motor, visual, and auditory. Ultimately, performance effectiveness when multitasking depends on the degree to which separate tasks use resources from the same sensory domain. Regardless of the application, users of passive BCI systems will likely perform activities involving multiple, disparate tasks engaging different sensory domains. Knowledge of which type of task(s) the user is performing at a given time, and the associated workload level, would allow for more appropriate adaptation schemes to be invoked by the BCI than would knowing just the overall level of workload (e.g., provide new information aurally when the user is already engaged in a task involving visual processing).

2.5. Relevant studies in the literature

The overall objective of this work is to develop algorithms to automatically identify, based on an individual's EEG signals, both the type and level of workload they are experiencing at a given time. In this work, the focus was on detecting workload in the visual and auditory sensory domains. To the best of our knowledge only one previous study has explored single-trial classification of tasks involving primarily visual and auditory processing via EEG. In their study, Putze et al. [11] showed that a "silent video watching" task and an "audiobook listening" task could be differentiated from each other (94% accuracy), and each from an idle state (91% accuracy for both) based on stimulus-locked EEG data (using either time- or frequency-domain features). While the results of this study are very promising, particularly given the high accuracies achieved, some questions remain to be explored. In their study, each single sensory perception task contained only stimuli from the target sensory domain, and the idle task contained no visual or auditory stimuli at all. It is unclear, then, how successful classification of visual and auditory processing would be in the presence of task-unrelated stimuli from the opposite domain. Or similarly, how successfully each sensory domain could be classified from an idle state that contains visual and auditory stimuli that the individual is passively experiencing but not attending to. These scenarios are more representative of a real-life scenario and need to be considered. Also, the study by Putze et al. [11] investigated visual and auditory perception tasks at just one level of task demand. It is not clear what effect the variability introduced by considering different workload levels, which again represents a more realistic scenario, would have on the ability to classify the type of sensory processing. These open questions are explored in this thesis.

Chapter 3: Methods

An original research experiment was designed and conducted to address the research questions of this thesis. Significant portions of this chapter were initially published in [12].

3.1. Study Participants

Six subjects (4 females and 2 male, mean age of 29 ± 2 years) participated in this study. The number of subjects recruited was limited due to Memorial University's abrupt suspension of research involving face-to-face interaction with human subjects resulting from the COVID-19 global pandemic in March 2020.

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

1) Were 18-65 years of age;

2) Had normal vision, or had vision that was corrected-to-normal with contact lenses;

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

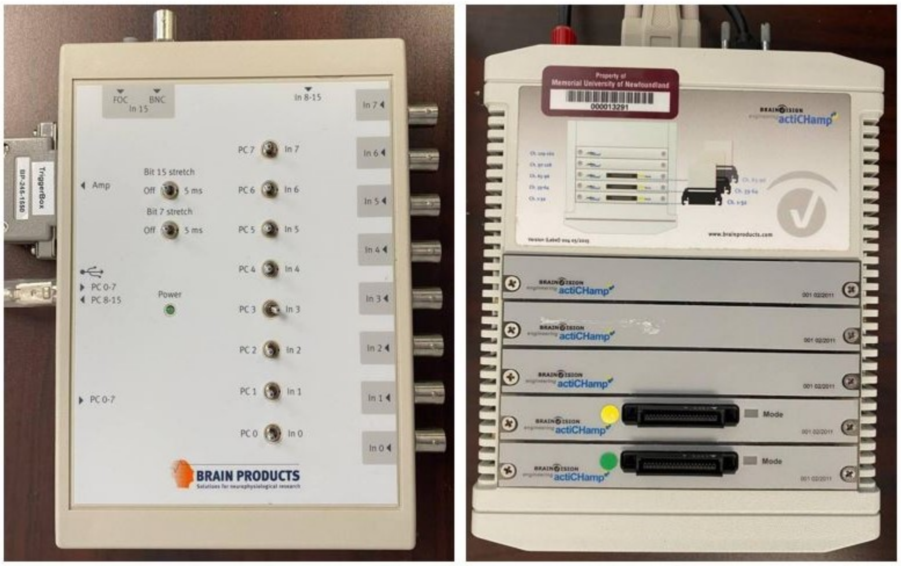
4) Had no history of neurological disorder, disease, or injury and no cognitive impairment.

In 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 as these activities can influence an individual's EEG signals. To maximize signal quality, participants were also asked to wash and dry their hair on the day of the experiment, and to refrain from using styling gel or other hair products, other than shampoo. Information regarding the participant's age, sex, and handedness was collected at the beginning of the session.

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

3.2. Instrumentation

EEG data were recorded using an ActiCHamp amplifier (Brain Products, GmBH) with 64-channel active electrodes positioned according to the "10-10 system" for electrode placement. Figure 5 shows the ActiChamp amplifier and triggerbox used in this study. Figure 6 shows the electrode cap used in this work. Electrolyte gel was used to achieve good coupling of each electrode to the scalp, the objective being to reduce the electrode impedance below 10 kΩ. The ground electrode was placed in the centre of the forehead, and electrode FCz was used as the reference electrode. EEG signals were recorded at a sampling rate of 500 Hz.



a) b)

Figure 5 a) Triggerbox b) ActiChamp Amplifier

Figure 6 Electrodes positioning and cap used in this work

3.3. Study Design

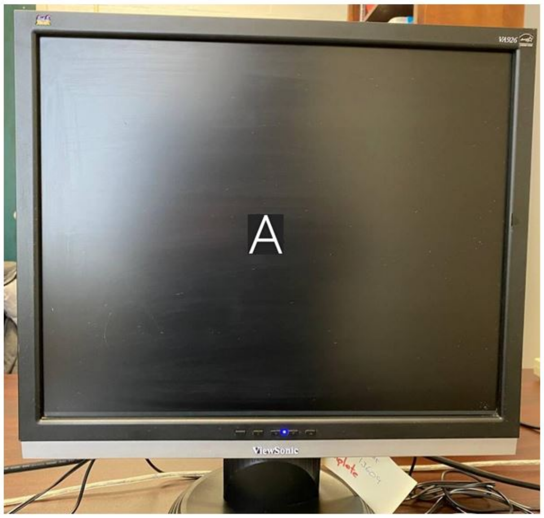
An experimental protocol was designed in which subjects performed two different versions of a task that differed based on the type of sensory processing involved (Auditory and Visual). Both versions of the task included stimuli from both sensory modalities, and differed only in the type of stimulus the subject was actively monitoring, or attending to. Each task was completed at two different levels of difficulty (Easy and Difficult). A dual task in which subjects had to attend to both the auditory and the visual stimuli was also completed under different workload conditions, but analysis of data from this task condition is beyond the scope of this thesis as these trials were designed to address a different set of research questions.

In selecting appropriate tasks, the goal was to ensure that 1) the tasks involved primarily auditory and/or visual processing, as appropriate, and required as little motor and cognitive workload as possible, 2) the workload of the task could be modulated with respect to the sensory processing required (and not in terms of some other sensory-independent cognitive component, like working memory), and 3) the auditory and visual tasks were as similar to one another as possible (within both workload levels) and differed only in terms of the type of sensory processing required. No suitable task paradigm could be found in the literature, therefore it was necessary to design an original, controlled laboratory task to meet these specific requirements. Specifically, a simple monitoring task was designed that had visual, auditory, and dual-task versions. The user interface for the experiment was implemented in the MATLAB Cogent2000 toolbox. Subjects completed the tasks using a desktop monitor, a standard keyboard, and standard computer speakers.

3.3.1. Description of experimental tasks

In each task trial, stimuli (letters A-Z from the English alphabet, and numbers 0-9) were presented in random order and the participant was instructed to respond to specified target stimuli by pressing the keyboard's space bar. The target stimuli were different for each trial, and the participant was told what they would be prior to starting the trial. Targets appeared at a rate of 30 ± 3%. The difficulty of the task was modulated by varying 1) the number of target letters, and 2) the stimulus presentation speed. In the Easy condition, there was just one target letter, and the stimuli were presented slowly (2250 ms from the start of one stimulus to the start of the next). In the Difficult condition, there were two target letters, and the stimuli were presented more quickly (750 ms from the start of one stimulus to the start of the next).

For Visual trials, the stimuli were presented via the computer monitor as white characters in the centre of a black screen (see Figure 7a). Stimuli were presented for a fixed interval (1500 ms for the Easy condition, 500 ms for the Difficult condition) followed by a fixed inter-stimulus interval with an all-black screen (750 ms for the Easy condition, 250 ms for the Difficult condition). For the Auditory trials, the stimuli were presented via speakers positioned on the desk in front of the subject.



a)

b)

Figure 7 a) A sample of English alphabet stimulus presented in this work b) A sample of Greek alphabet stimulus presented in this work

To ensure that any differences observed between the auditory and the visual trials were due to the sensory processing requirements of the monitoring task and not merely due to passive exposure to the stimuli, each task trial included a set of "passive stimuli" presented in the opposite sensory modality. This way, both visual and auditory monitoring trials included both visual and auditory stimuli, and they differed only in which type of stimuli the subject actually had to attend to/monitor during the trial. The passive stimuli were from the Greek alphabet, and changed at the same rate as the task stimuli for a given trial. Figure 7b shows a sample of a passive stimulus presented on a black screen. No target letters were given for the passive stimuli, and the subject was not required to respond to them in any way. Some letters from the Greek alphabet were excluded due to their visual similarity to English letters (e.g., kappa, κ). The Greek letters included in the set of passive stimuli were:

β (beta), Γ (gamma), γ (gamma), Δ (delta), ζ (zeta), η (eta), Θ (theta), θ (theta), Λ (lambda), λ (lambda), μ (mu), ξ (xi), Ξ (xi), Π (pi), π (pi), Σ (sigma), σ (sigma), Φ (phi), φ (phi), Ψ (psi), ψ (psi), Ω (omega), ω (omega), δ(delta). In trials that were not visual monitoring trials, subjects were instructed to keep their eyes open and look at the passive visual stimuli on the screen, but they were told not to pay attention to them or monitor them in any way. For trials that were not auditory monitoring trials, the passive auditory stimuli were played through the speakers the same way that actual task stimuli were in auditory monitoring trials; therefore subjects could not avoid being exposed to the passive auditory stimuli, but they were told not to pay attention or monitor them in any way.

Two different types of baseline trials, Baseline (with stimuli) and Baseline (without stimuli) were also recorded. For Baseline (with stimuli) trials, passive stimuli (from the Greek alphabet) were presented both visually and aurally. No targets were presented and the subject did not have to respond to either type of stimulus in any way. In half of these trials, the passive stimuli were presented at the speed of the "easy" task condition, and half at the speed of the "difficult" task condition. The baseline trials with no stimuli were "true baseline" trials in which the subjects were asked to simply focus their eyes on a constant "+" sign in center of the screen, and no auditory stimuli were presented.

Table 1 summarizes the different trial types.

Table 1 The types of trials completed during the experimental session © 2020 IEEE

Trial Type	Target Stimuli	Passive Stimuli	Difficulty	Description
Baseline (B)	No stimuli	No stimuli	N/A	The only stimuli presented is a constant "+" sign in the centre of screen on which subjects are asked to focus their eyes.
Baseline Easy (BE)	No stimuli	Visually and Aurally presented Greek alphabet	N/A	Subjects are played Greek alphabet stimuli during the trial and are asked to focus on the stimuli without any actions.
Baseline Difficult (BD)	No stimuli	Visually and Aurally presented Greek alphabet	N/A	Subjects are played Greek alphabet stimuli during the trial and are asked to focus on the stimuli without any actions.
Auditory Easy (AE)	Aurally presented English alphabet and numbers 0-9	Visually presented Greek alphabet	Easy – One target stimulus was presented with a fixed inter-stimulus interval of 750 ms	Subjects are played the target stimulus prior to starting the trial and are asked to attend to/monitor the auditory stimuli carefully and press the space key whenever they hear the target letters/numbers.
Auditory Difficult (AD)	Aurally presented English alphabet and numbers 0-9	Visually presented Greek alphabet	Difficult – Two target stimuli were presented with a fixed inter-stimulus interval of 250 ms	Subjects are played the target stimuli prior to starting the trial and are asked to attend to/monitor the auditory stimuli carefully and press the space key whenever they hear one of the target letters/numbers.
Visual Easy (VE)	Visually presented English alphabet and numbers 0-9	Aurally presented Greek alphabet	Easy – One target stimulus was presented with a fixed inter-stimulus interval of 750 ms	Subjects are played the target stimulus prior to starting the trial and are asked to attend to/monitor the visually stimuli carefully and press the space key whenever they hear the target letters/numbers.
Visual Difficult (VD)	Visually presented English alphabet and numbers 0-9	Aurally presented Greek alphabet	Difficult – Two target stimuli were presented with a fixed inter-stimulus interval of 250 ms	Subjects are played the target stimuli prior to starting the trial and are asked to attend to/monitor the visually stimuli carefully and press the space key whenever they hear one of the target letters/numbers.

3.3.2. Experimental Procedure

The experiment was completed in a single session of approximately three hours duration.

The protocol began with one eyes-closed and one eyes-open baseline trial (one minute each), followed by a practice block of six trials, including one Auditory, one Visual, two Auditory-Visual, and two Baseline (with stimuli) trials. Figure 8 shows timeline of experimental session and an example task block.

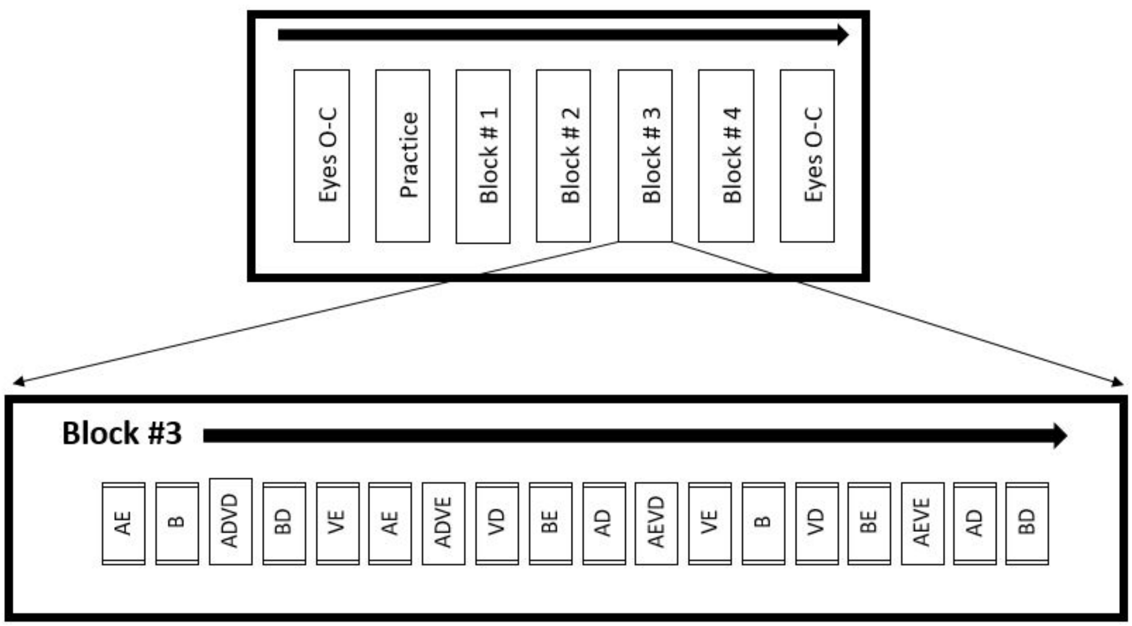


Figure 8 Timeline of experimental session and an example task block (Eyes O-C = block with one eyes-open and one eyes-closed baseline trial)

The main part of the experiment consisted of four blocks of 18 trials. Each trial was 30 seconds in duration. In each of the four blocks, there were two trials for each of the following seven conditions: Auditory-Easy (AE), Auditory-Difficult (AD), Visual-Easy (VE), Visual-Difficult (VD), Baseline-“Easy” (BE), Baseline-“Difficult” (BD), and Baseline (no stimuli) (B). There were also four dual task (i.e., Auditory-Visual) trials per block. The trial order was random in each block (except for the restriction that two trials of the same type would not appear back-to-back).

Thus, in total there were eight trials for each of the AE, AD, VE, VD, B, BE, and BD conditions, and 16 trials of the various Auditory-Visual conditions, for 72 trials in total.

Before each trial began, the subject was told what type of trial they were about to complete and reminded of the instructions for that trial type (via text explanation presented on the computer screen). For all non-Baseline trials, the subject was then presented with the target stimulus/stimuli for that trial. The targets were presented in the task modality of the particular trial (i.e., for Visual trials, the targets were presented visually on the screen while for Auditory trials, the targets were presented aurally via the speakers). The subject was able to repeat the target stimuli presentation as many times as needed prior to starting the trial. When ready, the subject started the trial by pressing the keyboard’s space bar. After each non-Baseline trial, subjects were asked to rate the level of mental effort required to perform the trial via a modified version of the Rating Scale of Mental Effort [77] (see Figure 9). This was chosen since it is quick and easy to complete as compared to other mental workload metrics, like the NASA-TLX. The subjects’ trial performance (response accuracy expressed as a percent), which served as an objective indicator of mental workload, was shown at the end of the trial. Figure 10 shows timeline of a single trial.

Participants progressed through trials at their own pace and could take breaks as needed between trials. They were also given breaks between blocks.

The experiment concluded with one eyes-closed and one eyes-open baseline trial (one minute each).

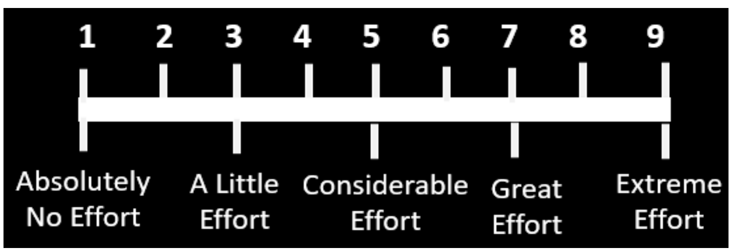


Figure 9 Rating Scale Mental Effort © 2020 IEEE

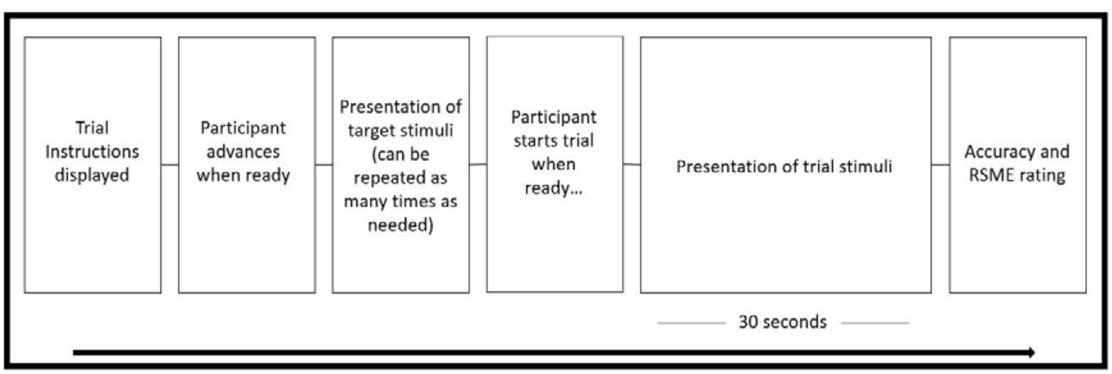


Figure 10 Timeline of a single trial

3.4. Data Analysis

3.4.1. Validation of induction of workload levels

In order to confirm that the Easy and Difficult conditions of the monitoring tasks induced different levels of workload, both subjective data (via the modified RSME ratings), and objective data (via the subjects' performance in the trials) were analyzed.

The participant's response accuracy was used as an objective task performance measure, and was calculated in each trial using the following formula:

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} * 100$$

where:

True Positive (TP) = Target stimuli presented and space key pressed

True Negative (TN) = No target stimuli presented and space key not pressed

False Positive (FP) = No target stimuli presented and space key pressed

False Negative (FN) = Target stimuli presented and space key not pressed

To compare workload levels, a two-way repeated measures ANOVA with two within-subjects factors, each with two levels (workload level: easy and difficult; task type: auditory and visual) was performed on both RSME rating and response accuracy.

3.4.2. EEG-based Task Classification

The main objectives of this work was to investigate 1) the ability to classify the type of task an individual is performing (in terms of required sensory processing, specifically auditory or visual) even when being exposed to passive, task-unrelated stimuli from the other sensory modality, 2) the effect of varying workload levels within each sensory modality on the ability to classify the modality, and 3) the ability to classify levels of workload within each sensory modality. This section details the classification analysis used to meet these objectives.

3.4.2.1. EEG pre-processing

EEG data tends to contain a lot of noise which can obscure weaker EEG signals. Artifacts such as blinking and muscle activity can contaminate the data. Pre-processing involved several steps to remove such artifacts.

The pipeline used was as follows: a band-pass filter from 1 Hz to 50 Hz was applied in order to remove the DC components of the signals, as well as signal frequencies beyond the range of interest. Then, segments contaminated with EMG and motion artifacts were manually rejected. Next, an independent component analysis (ICA) algorithm was applied using the EEGLAB toolbox in MATLAB, and components reflecting artifacts such as eye blinks and saccades were visually identified and removed. Figure 10 provides an example of four different EEG signal components produced using the ICA method. As can be seen from the topography in Figure 11, continuous data, and activity power spectrum graphs IC34, IC37, and IC45 are all examples of a bad components (i.e., ones that do not reflect neural activity) and should be removed. For IC28, which is a reliable component of a brain activity, it can be seen from the topography that there is a radial brain activity on the right side of the brain. Also brain activity is continuous through all the session. Also we have a smooth trend in power activity with some changes approximately below 10 Hz which indicates alpha activities. Following ICA-aided artifact removal, the data were down-sampled from 500 Hz to 256 Hz.

Figure 12 shows a sample multi-channel EEG signal segment a) before pre-processing, b) after EMG removal but before blink/saccade removal via ICA, and c) after all pre-processing steps, respectively.

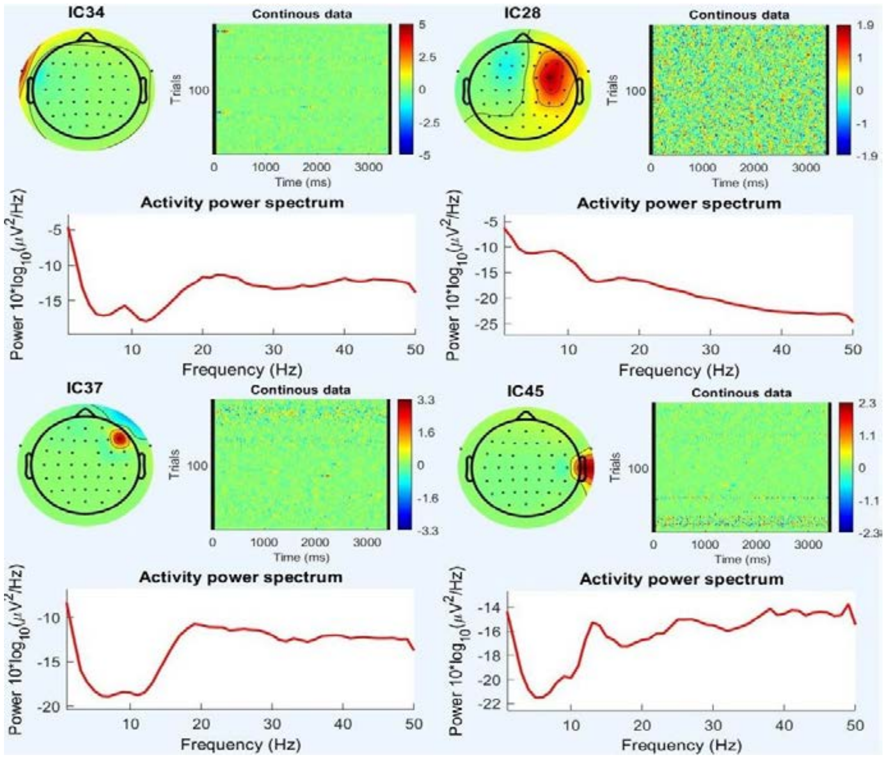


Figure 11 ICA Components of brain and non-brain activities

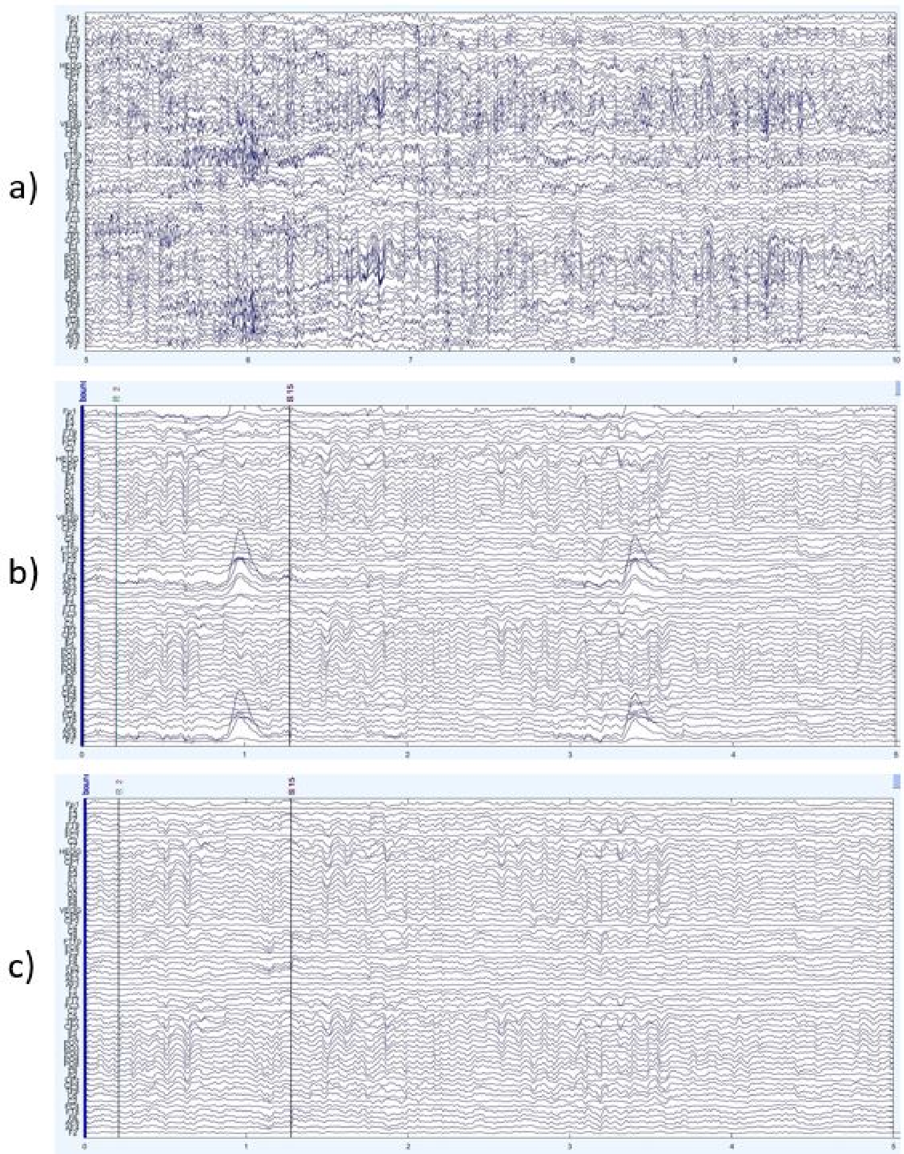


Figure 12 EEG signals a) before preprocessing, b) after EMG removal but before blink removal, and c) after all pre-processing steps

3.4.2.2. EEG signal feature calculation

Accuracy of learned models can be increased by extracting features from the raw input data. Here, frequency domain features of the EEG signals were calculated.

Because the motor requirements were slightly different for the Easy and Difficult task conditions (due to higher frequency of responding to stimuli in the latter case) the electrodes over the motor and sensorimotor brain regions, which included the 14 central and centro-parietal electrodes (Cz, C1-C6, CPz, CP1-CP6), were excluded for all Easy vs. Difficult classification problems as well as for all “task vs. baseline” classification problems. All other electrodes (49 in total) were included and considered individually in these cases. Since the motor requirements were matched in the Auditory and Visual task conditions (within a workload level) all 63 electrodes were considered for the Auditory vs. Visual classification problems.

For each electrode, signal power in seven common EEG frequency bands was calculated. Frequency bands were based on each subject’s individual alpha frequency (IAF) [78] as follows: Delta (IAF-8 to IAF-6), Theta (IAF-6 to IAF-4), Alpha1 (IAF-4 to IAF-2), Alpha2 (IAF-2 to IAF), Alpha3 (IAF to IAF+2), Beta (IAF+2 to IAF+20) and Gamma (IAF+20 to IAF+30). IAF was calculated as the average of the frequency of peak power within the frequency range of 5 to 15 Hz from the two eyes-closed baseline trials [28]. Also, signal power in non-overlapping frequency bands of width 1 Hz in the interval from 1 to 50 Hz was calculated (for a total of 49 frequency bands). Power time-series for each frequency band were obtained via the filter-Hilbert method. Then, the average power was calculated over non-overlapping one-second epochs.

Two scenarios were considered: 1) when the overall feature pool included all 56 frequency bands, and 2) when the overall feature pool included just the seven common EEG frequency bands. Thus, the total number of features considered in the first scenario was 2744 (i.e., 49 electrodes x 56 frequency bands) for the Easy vs Difficult and task vs. baseline classification problems, and 3528 (i.e., 63 electrodes x 56 frequency bands) for the Auditory vs. Visual classifications. The total number of features considered in the second scanrio was 343 (i.e., 49 electrodes x 7 frequency bands) for the Easy vs Difficult and task vs. baseline classification problems, and 441 (i.e., 63 electrodes x 7 frequency bands) for the Auditory vs. Visual classifications.

For each task condition (i.e., AE, AH, VE, VH, BE, BH, B), there was a total of 8 trials of 30 seconds duration recorded, which yielded approximately 240 samples/epochs per condition.

3.4.2.3. Feature selection

Feature selection is the process of reducing the dimensionality of data which tends to improve machine learning performance by removing redundant or ineffective features. In this work, the minimum redundancy maximum relevance (mRMR) feature selection approach was used. The mRMR is a feature selection approach that tends to select features with a high correlation with the class (output) and a low correlation between themselves. The mRMR algorithm ranks a set of features minimizing the redundancy among the subset of features while maximizing the relevance of the features to the classification problem at hand.

The effect of the feature set dimensionality on classification accuracy was explored. Feature set sizes between 5 and 50, in increments of 5, were considered.

3.4.2.4. Classification

Four different classification methods which are commonly used in EEG-based BCI research were explored: linear discriminant analysis without regularization (LDA), Regularized Linear discriminant analysis (rLDA), K-Nearest Neighbour (KNN), and Support Vector Machines (SVM). The performance of the classifiers was assessed via the classification accuracies obtained from 5 runs of 6-fold randomized cross-validation, performed separately for each subject. Features were selected for input to the classifiers (via the method described in the previous section) in each "fold" of the cross-validation based on the training data only. For the SVM classifier, all hyperparameters were optimized by cross-validation via the fitcsvm function in Matlab.

Several different binary classification problems were investigated to address the research questions of this thesis. First, to investigate if EEG could be effective for identifying the type of mental workload an individual is experiencing (research question #1, section 1.2.), Auditory vs. Visual monitoring trials were classified within each of the Easy and Difficult workload levels. To determine the effect of variation in mental workload on the ability to classify workload type (research question #2, section 1.2.), Auditory vs. Visual monitoring trials were classified with data from both workload levels combined. To determine if EEG could be effective for identifying distinct levels of mental workload within the visual and auditory domains (research question #3, section 1.2.), Easy vs. Difficult monitoring trials were classified within each task type.

Other classification problems that could be helpful in interpreting the results of those indicated above were also considered, including: 1) Easy vs. Difficult monitoring trials with data from both task types combined, and 2) each task condition (i.e., combination of mental workload type and level) vs. baseline (stimuli).

3.4.2.5. Statistical Analysis

To determine the effect of classifier type and feature pool (i.e., 7 or 56 frequency bands) on the classification accuracy, a two-way repeated measures ANOVA with two within-subjects factors was performed separately for the mental workload type (Auditory vs. Visual) classification problems and the mental workload level (Easy vs. Difficult) classification problems.

After selecting the best combination of classifier and feature pool, a one-way repeated measures ANOVA was performed to determine the effect of 1) mental workload level on the classification of mental workload type, and 2) mental workload type on the classification of mental workload level.

To determine if all classification accuracies found were greater than chance, the binomial test was used [79].

Chapter 4: Results

4.1. Validation of induction of workload levels

The average participant response accuracies for the monitoring trials for each workload type and workload level are given in Table 2. The repeated measures ANOVA revealed a significant effect of both workload type ($F_{(1,5)} = 9.51; p = .03$) and workload level ($F_{(1,5)} = 114.63; p < .001$) on response accuracy. There was no significant interaction effect ($F_{(1,5)} = 4.61; p = .09$). Post-hoc Tukey-Kramer tests revealed no significant difference in the response accuracies between the auditory and visual tasks within the easy condition ($t_{(5)} = 0.356; p = .75$), and a difference within the difficult condition that just reached significant ($t_{(5)} = 0.0134; p = .047$).

The average RSME ratings for each workload type and workload level are also given in Table 3. The repeated measures ANOVA revealed a significant effect of workload level ($F_{(1,5)} = 91.09; p < .001$) on RSME rating, but workload type had no significant effect ($F_{(1,5)} = 4.70; p = .08$). There was a significant interaction effect ($F_{(1,5)} = 12.08; p = .02$).

Table 2 Mean task performance ratings for each task type and workload level

Response Accuracy (%)			
	Auditory	Visual	Mean
Easy	99.09	100	99.56 ± 0.46
Difficult	94.72	98.07	96.40 ± 1.67
Mean	96.91 ± 2.18	99.04 ± 0.96	

Table 3 Mean RSME ratings for each task type and workload level

RSME Rating			
	Auditory	Visual	Mean
Easy	1.78	1.74	1.76 ± 0.02
Difficult	4.54	3.83	4.19 ± 0.35
Mean	3.16 ± 1.38	2.78 ± 1.04	

4.2. EEG-based state classification

Figure 13 shows the classification accuracies (averaged across all classification problems considered) for increasing feature set dimensionality. Based on visual analysis of this plot, as well as the amount of data available for the classification analysis (i.e., approximately 240 samples per class), results of the classification results in this section are based on using a 25-dimensional feature set chosen automatically via the mRMR feature selection algorithm. Results are reported for the four classifiers (i.e., LDA, rLDA, KNN, SVM) and two different feature pools (i.e., with 7 frequency bands and with 56 frequency bands) considered, as described in Sections 3.4.2.2. Only average accuracies (i.e., averaged across subjects) are reported here; see Appendix for detailed results for each participant, as well as for the results obtained using a smaller (specifically, 10-dimensional) feature set.

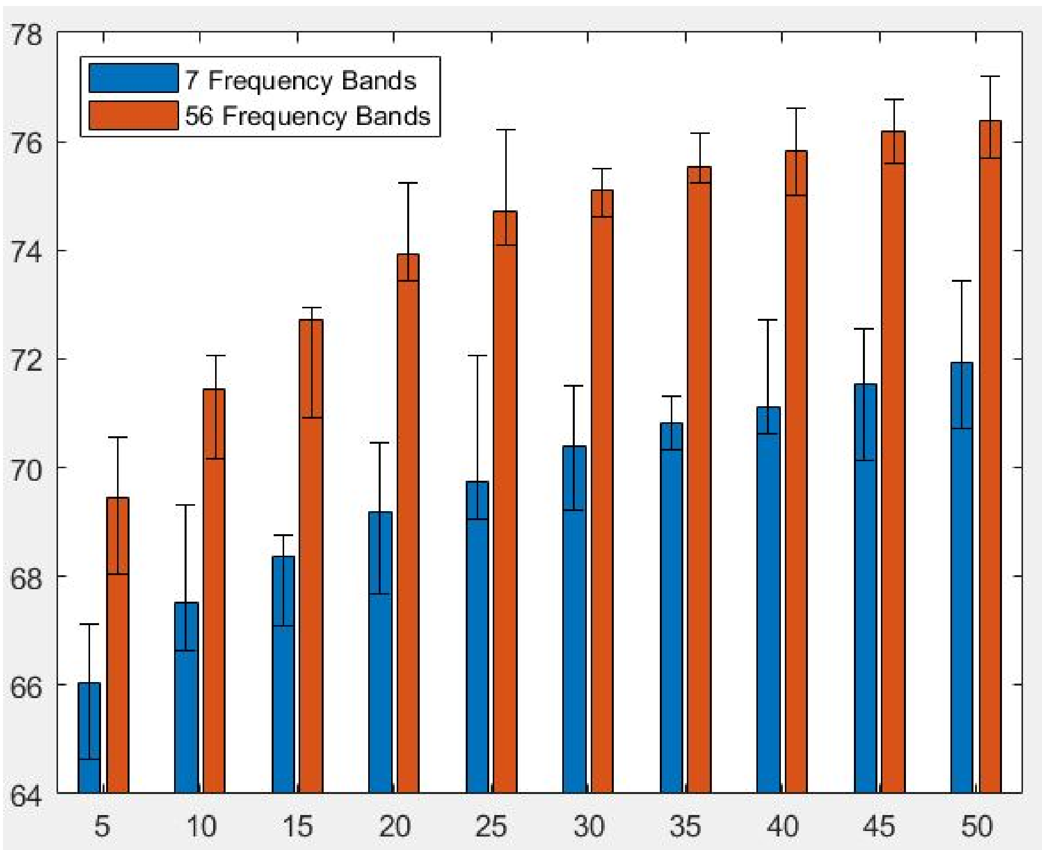


Figure 13 Classification accuracies (averaged across all classification problems considered) for different feature set sizes.

4.2.1. Classification of mental workload type: auditory vs. visual

Table 4 shows the results of the "workload type" classification problems (i.e., Auditory vs. Visual monitoring trials), both within a given workload level, and with both workload levels combined. Average accuracies (across participants) are given for each combination of classifier type and feature pool (i.e., with 7 or 56 frequency bands included) investigated. Repeated measure two-way ANOVA revealed a significant effect of classifier type ($F_{(3,15)} = 36.37; p < .001$) but no significant effect of feature pool ($F_{(1,5)} = 0.13; p = .73$) on classification accuracy. Post-hoc Tukey-Kramer tests indicated that there was no significant difference in classification accuracy between the LDA and rLDA classifiers ($t_{(15)} = 1.35; p = .20$) and that both of these classifiers had significantly greater accuracies than either the KNN or SVM classifiers ($t_{(15)} > 3.48; p < .004$). The SVM classifier achieved significantly higher accuracies than KNN ($t_{(15)} = 6.71; p < .001$).

Table 4 Across-subject average classification accuracies for the Auditory vs. Visual classification problems

Workload Type Classification: Auditory vs. Visual					
Classifier	Frequency Bands	Workload Level Included			
		Easy Only	Difficult Only	Both	Mean
LDA	7	64.7 ± 6.6	68.2 ± 4.5	63.9 ± 4.2	65.6
	56	68.1 ± 6.4	76.8 ± 4.9	68.9 ± 3.5	71.3
rLDA	7	65.0 ± 6.8	68.5 ± 4.2	63.5 ± 3.9	65.7
	56	67.5 ± 6.7	76.8 ± 5.4	68.8 ± 3.1	71.0

KNN	7	58.6 ± 6.4	62.7 ± 6.3	61.2 ± 4.8	60.8
	56	59.0 ± 3.2	67.9 ± 6.2	61.7 ± 3.4	62.9
SVM	7	64.9 ± 6.6	68.1 ± 5.4	65.9 ± 5.3	66.3
	56	61.0 ± 2.9	69.7 ± 6.4	63.3 ± 3.4	64.7

Based on the above results, the results from the rLDA classifier and the feature pool including all 56 frequency bands were taken for further investigation. Table 5 shows the workload type classification results, within each workload level and with workload levels combined, for each individual participant. By the binomial test, the upper limit of chance for n=240 samples/class (here there was one sample/1-second epoch x 30 seconds/trial x 8 trials/condition = 240 samples/class) and a significance level of α=0.05 is approximately 54.0%. For the cases where classification was done including Easy and Difficult trials together, the number of samples was doubled (n=480 samples/class), and in this case the upper limit of chance at a significance level of α=0.05 is approximately 52.7%. Based on the binomial test, all workload type classification accuracies - obtained for all participants, within each workload level and with workload levels combined - are significantly greater than chance.

The repeated measures one way ANOVA revealed a significant effect of "workload level included" on the workload type classification accuracies ($F_{(2,5)} = 13.98; p = .01$). Post-hoc Tukey-Kramer tests indicated that there was no significant difference between the "Easy only" and "Difficult only" conditions ($t_{(5)} = 2.79; p = .08$). Accuracies for the

"Difficult only" condition were significantly higher than for the "Both" condition ($t_{(5)} = 4.92; p = .01$) but accuracies for the "Easy only" condition were not ($t_{(5)} = 2.79; p = .08$).

Table 5 Per-participant accuracies for the Auditory vs. Visual classification problems. All accuracies were obtained with the regularized LDA (rLDA) classifier and a 25-dimensional feature set selected from the feature pool that included 56 frequency bands.

Workload Type Classification: Auditory vs. Visual				
Participant	Workload Level Included			Mean
	Easy Only	Difficult Only	Both	
1	62.1	72.7	64.9	66.6
2	75.9	85.1	74.1	78.4
3	71.7	67.9	67.7	69.1
4	69.8	77.4	67.5	71.6
5	55.6	77.3	66.7	66.5
6	69.8	80.3	71.7	73.9
Mean	67.5 ± 6.7	76.8 ± 5.4	68.8 ± 3.1	

4.2.2. Classification of workload level: easy vs. difficult

Table 6 shows the results of the "workload level" classification problems (i.e., Easy vs. Difficult monitoring trials), both within a given workload type, and with both workload types combined. Average accuracies (across participants) are given for each combination of classifier type and feature pool (i.e., with 7 or 56 frequency bands included) investigated.

Repeated measure two-way ANOVA revealed a significant effect of classifier type ($F_{(3,15)} = 27.65; p < .001$) but no significant effect of feature pool ($F_{(1,5)} = 1.68; p = .25$). Post-hoc Tukey-Kramer tests indicated that there was no significant difference in classification accuracy between the LDA and rLDA classifiers ($t_{(15)} = 2.19; p = .17$) but that both of these classifiers had significantly greater accuracies than both the KNN or SVM classifiers ($t_{(15)} > 3.48; p < .02$). The SVM classifier achieved significantly higher accuracies than the KNN classifier ($t_{(15)} = 3.26; p = .02$).

Table 6 Across-subject average classification accuracies for the Easy vs. Difficult classification problems.

Workload Level Classification: Easy vs. Difficult					
Classifier	Frequency Bands	Workload Type Included			
		Auditory Only	Visual Only	Both	Mean
LDA	7	74.7 ± 6.0	75.2 ± 2.5	69.3 ± 4.5	73.1
	56	79.4 ± 6.6	80.4 ± 5.5	72.5 ± 5.7	77.4
rLDA	7	74.8 ± 5.8	75.4 ± 2.8	68.9 ± 4.6	73.0
	56	81.1 ± 6.7	81.3 ± 6.3	72.8 ± 5.5	78.4
KNN	7	71.0 ± 7.3	66.3 ± 6.4	66.1 ± 5.0	67.8
	56	70.3 ± 8.4	73.0 ± 7.6	67.7 ± 7.3	70.3
SVM	7	75.7 ± 7.2	69.2 ± 5.3	68.5 ± 4.0	71.1
	56	71.9 ± 8.2	74.7 ± 7.7	68.7 ± 7.6	71.8

Based on the above results, the results from the rLDA classifier and the feature pool including all 56 frequency bands were again taken for further investigation. Table 7 shows the workload level classification results, within each workload type and with workload types combined, for each individual participant. Based on the binomial test, all workload level classification accuracies - obtained for all participants, within each workload type and with workload types combined - are significantly greater than chance.

The repeated measures one way ANOVA revealed a significant effect of "workload type included" on the workload level classification accuracies ($F_{(2,5)} = 78.63; p < .001$). Post-hoc Tukey-Kramer tests indicated that there was no significant difference between the "Auditory only" and "Visual only" conditions ($t_{(5)} = 1.05; p = .58$). Accuracies for both the "Auditory only" and "Visual only" conditions were significantly higher than for the "Both" condition ($t_{(5)} > 5.97; p < .0044$).

Table 7 Per-participant accuracies for the Easy vs. Difficult classification problems. All accuracies were obtained with the regularized LDA (rLDA) classifier and a 25-dimensional feature set selected from the feature pool that included 56 frequency bands.

Workload Level Classification: Easy vs. Difficult				
Participant	Workload Type Included			Mean
	Auditory Only	Visual Only	Both	
1	73.5	79.4	69.9	74.2
2	85.7	83.6	73.7	81.0
3	68.5	68.2	62.1	66.2

4	82.7	84.3	77.6	81.5
5	83.9	85.8	75.9	81.9
6	86.2	86.3	77.4	83.3
Mean	80.1 ± 6.7	81.3 ± 6.3	72.8 ± 5.5	

4.2.3. Classification of all task conditions vs. baseline

Table 8 shows the results of the classification of each individual task condition (i.e., all combinations of workload type and workload level) vs. baseline (with stimuli). In each case, the appropriate baseline condition is used; that is, that with the passive stimuli changing at the same speed as the task stimuli (e.g., Auditory Easy vs. Baseline "Easy").

Table 8 Across-subject average classification accuracies for all individual task conditions vs. corresponding baseline (with stimuli)

All Individual Task Conditions vs. Baseline					
Classifier	Frequency Bands	Task Condition			
		Auditory		Visual	
		Easy	Difficult	Easy	Difficult
LDA	7	70.3 ± 5.1	75.1 ± 7.8	68.4 ± 3.6	76.4 ± 5.0
	56	72.3 ± 4.6	79.4 ± 6.4	72.2 ± 3.2	81.7 ± 3.8
rLDA	7	70.2 ± 5.0	75.5 ± 8.0	68.3 ± 4.1	76.6 ± 5.1
	56	73.2 ± 5.3	79.9 ± 6.6	72.8 ± 4.9	82.2 ± 4.0
KNN	7	64.2 ± 5.0	71.2 ± 7.1	64.7 ± 4.0	71.4 ± 5.7

	56	62.6 ± 4.5	68.1 ± 9.5	64.1 ± 3.9	68.6 ± 6.2
SVM	7	69.4 ± 4.5	74.6 ± 7.1	68.4 ± 4.5	74.9 ± 7.3
	56	65.3 ± 3.3	71.8 ± 10.4	65.9 ± 2.7	69.9 ± 7.1

Based on the above results, the results from the rLDA classifier and the feature pool including 56 frequency bands were again taken for further investigation. Table 9 shows all task condition vs. baseline classification results, for each individual participant. Based on the binomial test, all of these classification accuracies are significantly greater than chance.

The repeated measures one way ANOVA revealed a significant effect of "workload level" on the task versus baseline classification accuracies ($F_{(1,5)} = 11.85; p = .02$), but no significant effect of "workload type" ($F_{(1,5)} = 0.73; p = .43$) and no interaction effect $\left(F_{(1,5)} = 4.72; p = .08\right)$.

Table 9 Per-participant accuracies for all individual task condition vs. baseline classification problems. All accuracies were obtained with rLDA classifier and a 25-dimensional feature set selected from the feature pool that included 56 frequency bands.

All Individual Task Conditions vs. Baseline				
	Task Condition			
	Auditory		Visual	
Participant	Easy	Difficult	Easy	Difficult
1	65.1	76.1	62.7	78.4
2	72.9	83.3	74.0	84.1

3	72.3	67.0	75.0	75.2
4	72.0	82.1	74.7	84.5
5	73.6	84.0	72.4	86.7
6	83.3	86.7	78.2	84.2
Mean	73.2 ± 5.3	79.9 ± 6.6	72.8 ± 4.9	82.2 ± 4.0

Chapter 5: Discussion

Overall, the RSME and task performance data suggests that the experimental protocol was effective in inducing different workload levels within each sensory modality since the repeated measures ANOVA revealed a significant effect of task difficulty on both measures. Regarding task type, the effect of this factor on task performance was significant, which is not desirable since ideally the difficulty of the tasks would be the same across modalities. However, since the response accuracy served as the task performance measure, it is likely that the difference between the modalities was not attributable to differences in task difficulty but was simply due to the fact that the stimuli could be perceived more quickly in the visual case than the auditory case, allowing the participant to respond more often within the given response interval in the visual trials. Post-hoc Tukey-Kramer tests indicated that there was no significant difference in response accuracy between the Visual and Auditory conditions within the Easy condition, and within the Difficult condition the difference just barely reached significance. There was no significant effect of task type on the RSME values, so it appears that the participants perceived the difficulty of both modalities to be the same. The results of these analyses suggest that our experimental protocol was effective.

The classification results suggest that trials involving primarily auditory and primarily visual processing can be distinguished at a level significantly exceeding chance based on EEG power spectral features even when the trials contain nearly identical overlapping stimuli from both sensory modalities, and differ only in the type of stimuli the individual is actually monitoring/attending to during the trial (research objective #1, section 1.2.). This seems to be true both when the required sensory processing is low (as in the Easy conditions) and high (as in the Difficult conditions). The average classification accuracy achieved in the Difficult condition was approximately 9% higher than in the Easy condition, but this difference was not statistically significant based on the repeated measures ANOVA and post-hoc Tukey tests. It would not be suprising if the Auditory and Visual conditions were indeed more separable in the Difficult condition when task-related activation would likely be stronger, and the lack of statistical significance could just be due to the relatively small sample size (n=6). When the data includes samples from conditions of both low and high processing requirement (i.e., both Easy and Difficult), the two types of sensory processing can still be distinguished at an accuracy greater than chance, however the results suggest that the accuracy may lower in this case. The accuracy was significantly lower when data came from both workload levels as compared to only one workload level, at least for the "Difficult Only" condition (research objective #2, section 1.2.). The results of workload level classification problems (i.e., Easy vs. Difficult monitoring trials), both within a given workload type, and with both workload types combined suggested that accuracies obtained for all participants were significantly greater than chance (research objective #3, section 1.2.)..

Unsurprisingly given the results discussed above, the auditory processing trials and visual processing trials could also be automatically distinguished from a baseline condition even when both the task trials and the baseline trials contain nearly identical overlapping stimuli of both types. As was observed for the classification of auditory and visual processing trials, this seems to be true both when the required sensory processing in the task trials is low (as in the Easy conditions) and high (as in the Difficult conditions), though as would be expected, accuracies are higher for the Difficult vs. Baseline case than the Easy vs. Baseline case.

The fact that the results for all classification problems, for all subjects, were significantly greater than chance is very encouraging. However, because of the small number of participants (n=6), the results of the classification analysis should be treated as preliminary and interpreted with caution. Data from more subjects are needed in order to get a reliable estimate of the differentiability of the investigated mental states, and to draw any conclusions about the differences among conditions (e.g., when auditory vs. visual processing classification is done within a workload level versus when both workload levels are included).

The classification results suggest that among the four different classifiers investigated, Linear Discriminant Analysis (LDA) and Regularized Linear Discriminant Analysis (rLDA) had higher performance in all classification problems. This is in line with the literature that suggests that linear classifiers have better performance in passive BCI investigations [47]. Based on the results, the performance of Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) classifiers also exceed the chance level and they are still promising through all classification problems.

Although most passive BCI studies that used signal power features have used just the seven common frequency bands, The combination of these seven bands with all 49 frequency bands of width equal to 1 Hz between 1-50 Hz has been investigated, and reached higher accuracies. As can be seen from Figure 12, the results in all the problems have slightly increased with using signal power extracted from all 56 frequency bands as compared to just seven. Overall from using 5 to 50 nominated final features, the trend of both graphs are very similar with the amount of accuracy increased in both these scenarios by approximately 6 percent. Statistical tests indicated that the difference in accuracy when using the 7 frequency bands vs. 56 frequency bands was not significant, however this is likely due to the small sample size (n=6), and that with more subjects this difference would become significant.

The classification accuracies achieved here are significantly lower than those reported in Putze et al. [11], the only other study identified that has attempted single trial classification of an auditory vs. a visual task using EEG. This could be at least in part because whereas Putze et al. used “pure” auditory, visual, and idle states, here the auditory and visual tasks were designed to be as similar as possible to one another (and to the baseline condition) in terms of the sensory stimuli, with the only difference between conditions being the type of stimuli the subject is actually attending to or monitoring. Future work will involve optimizing the algorithms to improve classification accuracy.

Chapter 6: Conclusions

6.1. Study conclusions

This work presented preliminary results of a study that aims to develop algorithms to detect both the type (specifically, auditory vs. visual processing) and level of mental workload experienced by an individual during task performance via EEG signals. Preliminary results are encouraging, suggesting that trials involving primarily auditory and primarily visual processing can be distinguished from one another, and individually from a baseline condition, even when the stimuli in all conditions are nearly identical and differ only in the type of stimuli the individual is actually attending to. Also, preliminary results suggesting that this algorithm can detect differences in workload level within a sensory modality as well. This work could result in the development of a passive BCI for mental workload detection that is able to determine both the level and type of workload being experienced by the user. With this capability, more appropriate environmental adaptation strategies could be invoked to improve the safety and/or performance of the operator being monitored.

6.2. Study limitations

The sample size for this study was relatively low, with only six subjects (4 females and 2 male, mean age of 29 ± 2 years) participating. The number of subjects recruited was limited due to Memorial University's abrupt suspension of research involving face-to-face interaction with human subjects resulting from the COVID-19 global pandemic in March 2020. As such, the results reported in this thesis, though promising, should be treated as preliminary until more data can be collected to verify their reliability.

6.3. Future work

Future work for this research will involve data collection from more participants in order to generalize the results and make them more reliable. Investigation and analysis of the dual task conditions (auditory and visual at the same time) will also be considered in future work. Further investigation of different EEG signal features and classification methods to to enhance and improve the results will also be taken into account.

References

[1] Brunner C, et al., "BNCI Horizon 2020: towards a roadmap for the BCI community," *Brain-Computer Interfaces,* vol. 2, no. 1, pp. 1-10, 2015.

[2] Zhang P, et al., "Spectral and temporal feature learning with two-stream neural networks for mental workload assessment," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 27, pp. 1149-1159, 2019.

[3] Salimi N, et al., "Mental workload classification using short duration EEG data: an ensemble approach based on individual channels," *IEEE Symposium Series on Computational Intelligence (SSCI)*, 2019.

[4] Dehais F, et al., "Monitoring pilot's mental workload using ERPs and spectral power with a six-dry-electrode EEG system in real flight conditions," *Sensors*, vol. 19, no. 6, 2019.

[5] Flumeri G, et al., "EEG-based mental workload assessment during real driving: A taxonomic tool for neuroergonomics in highly automated environments," *Neuroergonomics, Academic Press,* vol. 12, no. 509, pp. 121-126, 2019.

[6] Winnie K, et al., "An evaluation of mental workload with frontal EEG," *PLOS one,* vol. 12, no. 4, 2017.

[7] Yang S, et al., "Assessing cognitive mental workload via EEG signals and an ensemble deep learning classifier based on denoising autoencoders," *Computers in Biology and Medicine*, vol. 109, pp. 159-170, 2019.

[8] Bagheri M, et al., "EEG-based detection of mental workload level and stress: the effect of variation in each state on classification of the other," *Journal of Neural Engineering*, vol. 17, 2020.

[9] Aghajani H, et al., "Measuring Mental Workload with EEG+fNIRS," *Frontiers in Human Neuroscience*, vol. 38, 2017.

[10] Feltman K, et al., "Measuring the domain specificity of workload using EEG: Auditory and visual domains in Rotary-Wing simulated flight," *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 2020.

[11] Putze F, et al., "Hybrid fNIRS-EEG based classification of auditory and visual perception processes," *Frontiers in Neuroscience,* vol. 8, no. 373, 2014.

[12] Bagheri M, et al., "EEG-based classification of visual and auditory monitoring tasks © 20XX IEEE," *IEEE International Conference on Systems, Man, and Cybernetics (SMC),* pp. 4032-4037, 2020.

[13] Cattan G., "The Use of Brain–Computer Interfaces in Games Is Not Ready for the General Public," *Frontiers in computer science*, vol. 3, 2021.

[14] Yuanqing L, et al., "A Hybrid BCI System Combining P300 and SSVEP and Its Application to Wheelchair Control," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 11, pp. 3156-3166, 2013.

[15] Wodlinger B, et al., "Ten-dimensional anthropomorphic arm control in a human brain-machine interface: difficulties, solutions, and limitations," *Journal of Neural Engineering*, vol. 12, no. 1, 2014.

[16] Beyrouthy T, et al., "EEG Mind controlled Smart Prosthetic Arm," *IEEE International Conference on Emerging Technologies and Innovative Business Practices for the Transformation of Societies (EmergiTech)*, 2016.

[17] Zander T, et al., "Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general," *Journal of Neural Engineering*, vol. 8, no. 2, 2011.

[18] Lightbody G, et al., "The Brain Computer Interface: Barriers to Becoming Pervasive," in *Pervasive Health*, 2014, pp. 101-129.

[19] Aricò P, et al., "How Neurophysiological Measures Can be Used to Enhance the Evaluation of Remote Tower Solutions," *Frontiers in Human Neuroscience,* vol. 13, no. 303, 2019.

[20] Khan M, et al., "Passive BCI based on drowsiness detection: an fNIRS study," *Biomedical Optics Express,* vol. 6, no. 10, p. 4063–4078, 2015.

[21] Aricò P, et al., "Passive BCI beyond the lab: current trends and future directions," *Physiological Measurement*, vol. 39, no. 8, 2018.

[22] Krol L, et al., "Passive BCI-Based Neuroadaptive Systems," *Graz Brain-Computer Interface Conference*, 2017.

[23] Sheridan T, et al., "Toward the definition and measurement of the mental workload of transport pilots," 1979.

[24] Moray N., Mental Workload: Its Theory and Measurement, 1979.

[25] Hart S., "NASA-Task Load Index (NASA-TLX); 20 years later," *In Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, p. 904–908, 2006.

[26] Zijlstra F., Efficiency in work behaviour, Delft University, The Netherlands, 1993.

[27] Galy E, et al., "Measuring mental workload with the NASA-TLX needs to examine each dimension rather than relying on the global score: an example with driving," *Ergonomics*, vol. 61, no. 4, pp. 517-527, 2018.

[28] Bagheri M, et al., "EEG-based detection of mental workload level and stress: the effect of variation in each state on classification of the other," *Journal of Neural Engineering*, vol. 17, no. 5, 2020.

[29] Borghini G, et al., "Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness," *Neurosci Biobehav Rev,* vol. 44, pp. 58-75, 2014.

[30] Vogt J, et al., "The impact of workload on heart and blood pressure in en-route and tower air traffic control," *Journal of Psychophysiology,* vol. 20, no. 4, p. 297–314, 2006.

[31] Hancock P, et al., "Physiological reflection of mental workload," *Aviat Space Environ Med,* vol. 56, no. 11, pp. 1110-4, 1985.

[32] Brouwer A, et al., "Estimating workload using EEG spectral power and ERPs in the n-back task," *Journal of Neural Engineering,* vol. 9, no. 4, 2012.

[33] Rodriguez-Bermudez G, et al., "Testing Brain–Computer Interfaces with Airplane Pilots under New Motor Imagery Tasks," *International Journal of Computational Intelligence Systems,* pp. 937-946, 2019.

[34] Aricò P, et al., "A passive brain-computer interface application for the mental workload assessment on professional air traffic controllers during realistic air traffic control tasks," *Progress in Brain Research,* pp. 295-328, 2016.

[35] Stawicki P, et al., "Driving a Semiautonomous Mobile Robotic Car Controlled by an SSVEP-Based BCI," *Computational Intelligence and Neuroscience,* vol. 16, 2016.

[36] Yang Y, et al., "Toward brain-actuated car applications: Self-paced control with a motor imagery-based brain-computer interface," *Computers in Biology and Medicine,* vol. 77, pp. 148-155, 2016.

[37] G. D. Flumeri, et al., "EEG-based mental workload assessment during real driving: A taxonomic tool for neuroergonomics in highly automated environments," *Neuroergonomics,* pp. 121-126, 2019.

[38] Lakshmi M, et al., "Survey on EEG Signal Processing Methods," *International Journal of Advanced Research inComputer Science and Software Engineering,* vol. 4, no. 1, 2014.

[39] Bulent Usakli A., "Improvement of EEG Signal Acquisition: An Electrical Aspect for State of the Art of Front End," *Computational Intelligence and Neuroscience*, vol. 10, no. 2, 2010.

[40] Mantri S, et al., "Non invasive EEG signal processing framework for real time depression analysis," *SAI Intelligent Systems,* 2015.

[41] Singh S., "Magnetoencephalography: Basic principles," *Annals of Indian Academy of Neurology,* vol. 17, pp. 107-112, 2014.

[42] Abreu R, et al., "EEG-Informed fMRI: A Review of Data Analysis Methods," *Frontiers in human neuroscience,* vol. 12, no. 29, 2018.

[43] Blinowska K, et al., "Electroencephalography (EEG)," *Wiley Encyclopedia of Biomedical Engineering,* 2006.

[44] Laszlo S, et al., "A direct comparison of active and passive amplification electrodes in the same amplifier system," *Journal of Neuroscience Methods,* p. 298–307, 2014.

[45] Homan RW, et al., "Cerebral location of international 10–20 system electrode placement," *Electroencephalography and Clinical Neurophysiology,* vol. 66, no. 4, pp. 376-382, 1987.

[46] Teplan M., "Fundamental of EEG Measurement," *Measurement Science Review*, vol. 2, no. 2, 2002.

[47] Lotte F, et al., "A review of classification algorithms for EEG-based brain computer interfaces: a 10 year update," *Journal of Neural Engineering,* vol. 15, no. 3, 2018.

[48] Jasper H, et al., "Human Brain Rhythms: I. Recording Techniques and Preliminary Results," *The Journal of General Psychology,* pp. 98-126, 1936.

[49] Aparnathi R, et al., "The study About Brain Wave Extreme Low frequency and Works," *International medical Association Expert talk & Conference (India),* vol. 1, 2014.

[50] Ding C, et al., "Minimum Redundancy Feature Selection from Microarry Gene Expression Data," *Journal of Bioinformatics and Computational Biology,* vol. 3, pp. 185-205, 2005.

[51] Hejazi M, et al., "Input variable selection for water resources systems using a modified minimum redundancy maximum relevance (mRMR) algorithm," *Advance in Water Resources*, vol. 32, no. 4, pp. 582-593, 2009.

[52] James G, et al., An introduction to statistical learning, California: Springer, 2013.

[53] Fukunaga K., Introduction to Statistical pattern recognition, Academic Press Inc., 2nd Edition 1990.

[54] Khan MJ, et al., "Drowsiness detection using fNIRS in different time windows for a passive BCI," *IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob),* 2016.

[55] Khan MJ, et al., "Drowsiness detection in dorsolateral-prefrontal cortex using fNIRS for a passive-BCI," *International Conference on Control, Automation and Systems (ICCAS),* vol. 15, 2015.

[56] Garrett D, et al., "Comparison of linear, nonlinear, and feature selection methods for eeg signal classification," *IEEE Transactions on neural systems and rehabilitation engineering,* vol. 11, no. 2, pp. 141-144, 2003.

[57] Friedman J., "Regularized Discriminant Analysis," *Journal of the American Statistical Association*, vol. 18, no. 405, pp. 165-175, 1989.

[58] Shuiwang J, et al., "Kernel Uncorrelated and Regularized Discriminant Analysis: A Theoretical and Computational Study," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 10, pp. 1311 - 1321, 2008.

[59] Maggi L, et al., "A four command BCI system based on the SSVEP protocol," *International Conference of the IEEE Engineering in Medicine and Biology Society*, p. 1264–1267, 2006.

[60] Mak J, et al., "Optimizing the P300-based brain–computer interface: current status, limitations and futuredirections," *Journal of Neural Engineering*, 2011.

[61] Rashid M, et al., "The Classification of EEG Signal Using Different Machine Learning Techniques for BCI Application," *International Conference on Robot Intelligence Technology and Applications*, pp. 207-221, 2019.

[62] Burges C., "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery*, pp. 121-167, 1998.

[63] Kaur B, et al., "EEG Based Emotion Classification Mechanism in BCI," *Procedia Computer Science*, vol. 132, pp. 752-758, 2018.

[64] Islam F., "A robust and reliable online P300-based BCI system using Emotiv EPOC + headset," *Journal of Medical Engineering & Technology,* vol. 45, pp. 94-114, 2021.

[65] Ullah Amin H, et al., "Classification of EEG Signals Based on Pattern Recognition Approach," *Frontiers in Computational Neuroscience,* vol. 11, no. 103, 2017.

[66] Jiawei H, et al., Data Mining: Concepts and Techniques, University of Illinois: Elseivier, Third Edition 2012.

[67] P. Zhang, et al., "Spectral and temporal feature learning with two-stream neural networks for mental workload assessment," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* pp. 1149-1159, 2019.

[68] N. Salimi, et al., "Mental workload classification using short duration EEG data: an ensemble approach based on individual channels," *IEEE Symposium Series on Computational Intelligence (SSCI),* 2019.

[69] Fazli S, et al.,, "Enhanced performance by a hybrid NIRS–EEG brain computer interface," *Neuroimage,* vol. 59, no. 1, pp. 519-529, 2012.

[70] Tomita Y, et al., "Bimodal BCI using simultaneously NIRS and EEG," *IEEE Transactions on Biomedical Engineering,* vol. 61, no. 4, pp. 1274-1284, 2014.

[71] K. A. Feltman, et al., "Measuring the domain specificity of workload using EEG: Auditory and visual domains in Rotary-Wing simulated flight," *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 2020.

[72] Wickens C., "Multiple resources and mental workload," *Human Factors: The Journal of the Human Factors and Ergonomics Society,* p. 449–455, 2008.

[73] Basil M., "Multiple Resource Theory," Encyclopedia of the Sciences of Learning, [Online]. Available: https://link.springer.com/referenceworkentry/10.1007%2F978-1-4419-1428-6_25.

[74] Wickens C., "The structure of attentional resources," *Attention and performance VIII,* p. 239–257, 1980.

[75] Wickens C., "Processing resource in attention," *Parasuraman R, Davies R, Varieties of attention,* p. 63–101, 1984.

[76] Wickens C., "Situation awareness and workload in aviation," *Current Directions in Psychological Science,* p. 128–133, 2002.

[77] Alimohammadi I, et al.,, "Validity and Reliability of Rating Scale Mental Effort, Integrated Workload Scale, and Overall Workload Scale in Iran," *International Journal of Occupational Hygiene*, vol. 11, 2019.

[78] Klimesch W., "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.

[79] Müller-Putz G, et al., "Better than random? A closer look on BCI results," *International Journal of Bioelectromagnetism,* vol. 10, no. 1, pp. 52-55, 2008.

[80] Cain B., "A review of the mental workload literature," Defence Research and Development, Toronto, Canada, 2007.

Appendices

	P1	P2	P3	P4	P5	P6
AEvsVE	57.5000	70.7083	67.3577	59.2308	53.9167	62.2500
ADvsVD	66.0569	73.4553	61.2393	69.4715	62.5203	67.0325
AvsV(All)	61.7300	68.6214	62.0833	59.1770	56.4815	64.1975
AEvsAD	70.1709	77.7083	62.5641	70.8547	72.6250	81.7500
VEvsVD	70.9649	72.0833	68.0741	75.8750	69.2083	77.0000
EvsD(All)	65.9559	65.9592	59.0525	70.3477	67.9167	74.5595
AEvsBE	61.0256	68.3761	65.9167	68.2906	72.1250	75.0417
ADvsBD	66.9048	73.8889	54.9145	78.3730	78.2937	79.2460
VEvsBE	61.5789	63.5043	61.5833	72.7578	70.7083	70.2917
VDvsBD	67.2358	72.7236	69.7083	78.2114	78.6585	74.3496
BEvsBD	62.2500	61.6667	66.5417	69.6250	62.5000	72.5417

LDA Classifier using 7 Frequency band features with 10 best Features selected

	P1	P2	P3	P4	P5	P6
AEvsVE	61.9298	77.9583	72.7236	66.5385	58.9583	64.7083
ADvsVD	68.0081	78.9431	63.3761	71.7073	63.9431	71.9106
AvsV(All)	64.0928	72.2222	65.6458	62.4280	58.7449	66.0905
AEvsAD	73.5470	81.9167	67.7350	74.9573	74.1250	84.9583

VEvsVD	76.7544	72.2917	70.8148	79.6250	76.7500	78.2917
EvsD(All)	69.1422	68.7890	61.6895	72.7578	70.9881	75.3095
AEvsBE	63.9316	72.6068	67.7083	71.5812	73.2083	79.1250
ADvsBD	70.0397	77.9762	61.7094	80.4762	82.9762	83.2143
VEvsBE	65.8333	68.4615	64.3750	71.5833	71.8333	73.6667
VDvsBD	69.7967	79.2276	72.9583	83.1301	82.0325	73.2520
BEvsBD	65.5417	65.8547	67.5000	70.3750	65.4167	77.8333

LDA Classifier using 7 Frequency band features with 35 best Features selected

	P1	P2	P3	P4	P5	P6
AEvsVE	60.7456	74.4167	69.3902	64.7436	55.0417	65.3333
ADvsVD	70.4878	83.0081	62.6923	75.7317	74.8780	77.3577
AvsV(All)	62.7637	72.8189	63.2917	66.8519	63.0658	69.7531
AEvsAD	72.9060	83.7083	62.9915	78.9744	78.5417	77.8750
VEvsVD	74.2105	77.0417	64.2963	81.2500	80.9583	79.7500
EvsD(All)	66.5686	70.8153	59.6119	75.0959	71.6190	74.5119
AEvsBE	61.8376	66.3248	66.9167	67.0085	71.9167	78.3750
ADvsBD	72.9365	78.2143	61.2821	77.8968	79.3254	81.9841
VEvsBE	61.8860	73.5043	68.1250	68.8333	70.9583	70.8333
VDvsBD	75.0000	77.6016	73.4167	79.3089	83.8618	78.0081

BEvsBD	68.3333	67.7350	73.2917	70.7500	64.9167	73.4167

LDA Classifier using 56 Frequency band features with 10 best Features selected

	P1	P2	P3	P4	P5	P6
AEvsVE	63.2018	77.0417	75.1220	71.7094	55.5833	70.9583
ADvsVD	74.2683	85.8943	69.4872	77.9675	77.3577	80.3659
AvsV(All)	66.3713	75.3292	68.2708	68.7449	65.0823	73.6214
AEvsAD	75.5983	86.8750	68.8034	81.8803	84.7500	87.1667
VEvsVD	78.5965	84.0000	72.2963	86.0417	85.4583	85.7917
EvsD(All)	70.6250	74.4245	61.5525	79.2686	76.4524	77.6905
AEvsBE	66.7094	73.6325	70.8750	70.9402	75.9583	80.0000
ADvsBD	75.0397	83.7698	67.8632	81.9444	86.0317	87.5794
VEvsBE	64.7368	74.2308	77.8333	75.1667	74.7917	71.7500
VDvsBD	79.7967	85.4878	75.5417	84.3089	87.3171	82.6829
BEvsBD	71.9583	70.3419	75.2917	74.0000	69.7917	79.5417

LDA Classifier using 56 Frequency band features with 35 best Features selected

	P1	P2	P3	P4	P5	P6
AEvsVE	59.0789	70.6250	67.4797	59.2735	54.8333	62.9167
ADvsVD	66.1382	73.1301	59.6581	68.0894	62.2764	67.4390
AvsV(All)	61.8354	69.0123	62.0208	58.9918	56.4403	64.9794

AEvsAD	70.9402	77.8750	64.1042	71.8803	73.0417	81.5417
VEvsVD	71.7544	72.5833	68.4074	76.1250	69.1250	77.0833
EvsD(All)	66.4828	66.1151	59.1781	70.5995	68.2857	74.6190
AEvsBE	61.5385	68.0342	64.9167	67.9060	71.7083	75.3333
ADvsBD	66.6270	74.3254	55.6410	77.7778	78.9286	79.8810
VEvsBE	60.6140	63.5470	60.9583	70.2917	71.7917	70.9167
VDvsBD	68.9431	72.5610	70.2083	79.2276	79.0244	74.8780
BEvsBD	63.2917	61.8803	66.0000	70.1667	61.0833	71.0417

rLDA Classifier using 7 Frequency band features with 10 best Features selected

	P1	P2	P3	P4	P5	P6
AEvsVE	61.1842	78.2083	72.0732	67.2650	57.7500	65.3750
ADvsVD	68.5772	78.6585	64.5299	70.4472	64.4309	71.9919
AvsV(All)	63.7342	71.8313	64.1042	62.2222	58.9712	65.8025
AEvsAD	73.1197	80.4167	66.7094	75.0855	75.6250	85.0833
VEvsVD	77.7632	73.2083	70.0370	80.4583	75.9583	77.5417
EvsD(All)	68.8603	69.4844	60.9589	72.9376	70.6905	75.5357
AEvsBE	64.8718	71.1966	67.6667	70.9402	73.7917	79.3333
ADvsBD	70.8730	78.1746	61.1538	81.0317	82.8571	82.5794
VEvsBE	66.2281	68.1624	63.0417	71.6667	72.6250	72.7500

VDvsBD	70.0407	78.1707	73.3750	83.8211	82.4390	75.2439
BEvsBD	64.2083	66.7949	66.2917	71.0417	63.8333	76.0833

rLDA Classifier using 7 Frequency band features with 35 best Features selected

	P1	P2	P3	P4	P5	P6
AEvsVE	60.3509	75.0000	69.7154	64.3162	56.0417	66.2500
ADvsVD	70.1626	83.5772	64.0171	75.7724	75.0407	78.7398
AvsV(All)	62.5527	72.4691	63.0208	67.0782	65.1029	68.7654
AEvsAD	72.9487	83.7083	62.1795	79.5299	77.9167	82.7917
VEvsVD	77.1930	78.3750	64.4444	81.6667	80.8333	81.9583
EvsD(All)	66.0784	71.3909	59.4635	75.3477	72.8452	76.1667
AEvsBE	62.8205	68.0769	69.9583	71.1111	70.0833	79.0000
ADvsBD	74.1270	79.8413	62.8205	79.3651	79.8810	81.8651
VEvsBE	61.5789	72.9915	68.5417	73.1667	71.0417	75.5000
VDvsBD	76.0976	78.8618	73.7083	82.3577	83.8618	80.5285
BEvsBD	67.0833	66.2821	73.4583	71.2917	66.0833	73.0833

rLDA Classifier using 56 Frequency band features with 10 best Features selected

	P1	P2	P3	P4	P5	P6
AEvsVE	64.0789	77.0417	75.1220	70.8120	56.7500	72.0833

ADvsVD	74.2683	86.4634	70.2137	77.7236	78.4553	80.8537
AvsV(All)	66.2236	74.6708	68.7917	68.8066	66.6255	72.5514
AEvsAD	75.7265	86.8750	70.6838	83.7179	85.2500	88.0000
VEvsVD	79.8684	84.6250	70.8148	84.5833	86.6667	86.6667
EvsD(All)	71.2377	74.7602	62.4201	79.4604	77.3214	78.3214
AEvsBE	67.4359	74.8291	73.3750	73.8889	74.5000	82.8333
ADvsBD	77.6587	84.7222	67.6496	83.7698	86.1905	88.3333
VEvsBE	63.2018	73.6752	77.2917	75.6667	72.4583	78.1250
VDvsBD	80.5285	85.6504	75.4583	86.3008	87.5203	85.4472
BEvsBD	72.0000	70.7265	75.1667	74.5833	72.0000	79.2500

rLDA Classifier using 56 Frequency band features with 35 best Features selected

	P1	P2	P3	P4	P5	P6
AEvsVE	53.1579	71.9583	64.1057	58.5897	56.4583	58.5000
ADvsVD	64.5528	74.1057	57.1795	62.1545	60.0407	64.5122
AvsV(All)	60.0844	69.0741	65.4375	57.2840	57.2428	63.5185
AEvsAD	66.8376	74.4167	61.8376	67.8632	69.7500	83.6250
VEvsVD	62.5877	65.4167	60.2963	65.8333	63.4583	77.7917
EvsD(All)	61.6176	61.7746	59.9772	66.8585	65.2024	75.2143
AEvsBE	60.0855	63.0342	66.4167	65.2991	68.3333	70.5417

ADvsBD	65.8730	73.3730	56.8376	68.2143	78.4127	76.5079
VEvsBE	57.4123	65.9829	60.3750	62.8750	69.0833	67.0000
VDvsBD	61.2195	74.4309	66.5417	69.2276	75.1626	74.2683
BEvsBD	59.4167	60.8547	65.0833	68.3750	62.5833	68.7500

KNN Classifier using 7 Frequency band features with 10 best Features selected

	P1	P2	P3	P4	P5	P6
AEvsVE	48.1579	71.1250	61.9106	52.6068	52.5000	58.8333
ADvsVD	59.5122	73.4553	56.3248	60.1626	61.2195	66.2602
AvsV(All)	60.1899	70.2675	57.5625	57.2840	56.5638	63.0453
AEvsAD	67.6068	77.9167	59.9573	68.0342	69.4583	83.2917
VEvsVD	59.8246	64.0417	60.1481	67.2500	67.8333	80.0833
EvsD(All)	62.9657	65.2518	59.7260	68.0815	66.0595	75.6548
AEvsBE	60.0000	60.0000	58.4583	64.9573	71.8333	69.7500
ADvsBD	60.8730	73.1349	60.3419	72.7381	79.4444	77.5794
VEvsBE	58.9035	65.8547	61.1667	63.6250	70.4583	69.0417
VDvsBD	60.7724	73.9837	64.9167	72.2764	77.6016	74.8374
BEvsBD	62.6667	64.0171	64.2917	67.6667	61.1667	70.2500

KNN Classifier using 7 Frequency band features with 35 best Features selected

	P1	P2	P3	P4	P5	P6
AEvsVE	57.8947	60.8750	59.2276	56.6667	52.9167	60.0833
ADvsVD	64.6748	66.3821	58.3333	70.6911	72.8862	72.7642
AvsV(All)	60.1266	59.3621	54.3333	65.6379	61.0905	64.9588
AEvsAD	65.1709	68.4583	52.4359	77.9487	70.3750	79.3333
VEvsVD	67.2807	67.1250	59.6667	75.3333	79.7083	78.9583
EvsD(All)	61.9363	63.8010	53.0936	71.4508	69.7738	75.5833
AEvsBE	59.8718	59.9145	59.5833	64.6154	61.5833	68.2083
ADvsBD	61.3492	58.8095	56.1111	69.6429	78.6905	78.1746
VEvsBE	56.4474	57.7350	63.0833	69.2083	63.7917	65.7917
VDvsBD	63.2520	63.9431	58.2917	73.7805	71.4228	74.9187
BEvsBD	59.1250	59.1880	70.7083	65.3750	61.7500	70.8750

KNN Classifier using 56 Frequency band features with 10 best Features selected

	P1	P2	P3	P4	P5	P6
AEvsVE	58.1579	63.8750	58.5366	61.9658	53.7083	61.0000
ADvsVD	60.1626	70.6911	58.8889	71.5447	74.1870	74.8374
AvsV(All)	60.1055	64.2387	57.0417	66.9959	63.2716	66.7078
AEvsAD	64.3590	71.2083	56.4103	77.5214	72.2083	82.9583
VEvsVD	67.8509	71.4583	61.3333	77.3333	80.7917	82.2500
EvsD(All)	64.2647	67.3981	54.8402	73.7290	72.7857	78.0833

AEvsBE	62.2222	59.3590	60.2500	66.1111	65.5417	70.2917
ADvsBD	58.0159	66.4286	55.8974	72.7778	79.0873	80.1190
VEvsBE	58.9035	61.5812	68.7083	64.9583	67.2083	66.9583
VDvsBD	61.2602	66.4634	62.4583	76.3821	71.4228	78.7805
BEvsBD	58.8333	64.1026	69.4167	69.7917	64.8750	70.8750

KNN Classifier using 56 Frequency band features with 35 best Features selected

	P1	P2	P3	P4	P5	P6
AEvsVE	55.5263	74.8750	68.2927	59.2308	57.1250	59.5000
ADvsVD	64.5935	73.4553	62.7778	63.8211	63.4959	65.9756
AvsV(All)	60.2954	70.7613	70.2917	59.0329	59.1564	64.9177
AEvsAD	66.8803	79.4167	69.5726	66.5385	71.8333	81.9167
VEvsVD	65.0000	65.3750	63.0741	64.6250	64.6667	75.5417
EvsD(All)	61.0417	69.6581	60.9247	65.2878	67.5000	74.0595
AEvsBE	59.4444	68.0769	70.7917	65.9402	71.2083	74.6667
ADvsBD	64.9206	77.8175	58.7607	72.0238	79.8413	78.0159
VEvsBE	61.2281	68.5897	59.4167	65.1250	71.0000	68.9583
VDvsBD	57.3984	74.1870	69.0417	71.6667	78.4146	76.5041
BEvsBD	59.7500	64.2308	66.8750	67.7500	64.7917	70.9583

SVM Classifier using 7 Frequency band features with 10 best Features selected

	P1	P2	P3	P4	P5	P6
AEvsVE	51.0088	78.5417	70.5285	65.8120	59.0833	61.8750
ADvsVD	63.8211	78.2520	62.6923	66.3008	69.6341	72.5203
AvsV(All)	62.3418	75.5556	65.8333	62.4691	61.4815	67.0165
AEvsAD	69.7863	86.0000	69.7009	69.6154	73.6667	86.1667
VEvsVD	62.2368	70.6667	64.7037	71.1250	74.1667	78.5833
EvsD(All)	65.0368	70.2518	65.5822	68.9568	68.9286	76.8690
AEvsBE	62.6068	69.4872	67.0833	69.6154	76.5833	73.9583
ADvsBD	64.3254	77.7778	64.5299	76.6667	81.5079	83.6905
VEvsBE	62.9386	74.1453	65.2917	67.0417	75.9167	70.7500
VDvsBD	61.5447	80.5285	69.9167	79.0650	83.3740	78.9837
BEvsBD	62.9583	69.6581	69.9583	70.4167	67.5000	73.5833

SVM Classifier using 7 Frequency band features with 35 best Features selected

	P1	P2	P3	P4	P5	P6
AEvsVE	57.7632	61.2917	58.6179	55.8974	56.5000	59.1667
ADvsVD	64.4309	63.4146	56.4957	70.0813	72.4390	71.5854
AvsV(All)	59.7679	57.7160	55.9167	63.6214	60.8642	63.4156
AEvsAD	62.6923	69.2083	54.7009	76.6239	69.6667	77.1667
VEvsVD	64.5614	68.6667	58.4074	73.6667	78.3333	77.5417

EvsD(All)	60.5760	62.9137	53.0023	71.8825	69.4167	74.8214
AEvsBE	59.9145	60.7692	58.4167	62.5641	64.4583	66.7083
ADvsBD	58.3333	59.5238	59.7436	70.7143	81.0714	80.6746
VEvsBE	59.8246	59.0598	62.5000	65.5833	62.7500	67.0833
VDvsBD	63.0488	62.2764	58.1250	74.4309	69.3902	75.4878
BEvsBD	56.9167	59.7436	66.6250	65.1667	61.2917	68.4167

SVM Classifier using 56 Frequency band features with 10 best Features selected

	P1	P2	P3	P4	P5	P6
AEvsVE	60.5702	66.5833	63.3740	61.0684	56.8333	62.3333
ADvsVD	63.4959	71.5447	60.1282	75.6504	76.9512	78.3740
AvsV(All)	61.9409	65.3086	59.5417	68.2922	65.2263	68.8683
AEvsAD	65.2564	75.6250	58.7607	80.3419	76.4167	80.9583
VEvsVD	71.4912	76.9167	62.8519	80.8333	81.5833	84.6250
EvsD(All)	65.4657	69.7482	57.1918	75.6715	73.7381	79.2381
AEvsBE	65.8547	65.8547	62.0000	67.6923	69.3750	73.2500
ADvsBD	57.3016	68.4524	62.7778	76.7063	84.8810	84.8016
VEvsBE	63.6404	64.7009	70.6250	67.6667	69.9167	68.2083
VDvsBD	60.8943	69.6748	64.1250	77.8862	73.5772	82.1951
BEvsBD	60.0417	68.3333	71.7917	71.0417	68.5833	72.7083

SVM Classifier using 56 Frequency band features with 35 best Features selected