Cognitive Task Difficulty Analysis using EEG and Data Mining

Aruna Duraisingam

September 2016

Dissertation submitted in partial fulfilment for the degree of

Master of Science in Advanced Computer Science

(Computational Intelligence)

School of Computing

University of Kent

Canterbury, Kent

Word Count - 12000

Abstract

The aim of this research is to measure cognitive load based on different task difficulty level (TDL) for program comprehension tasks using Electroencephalogram (EEG) and classify using machine learning techniques. Existing research on task difficulty and program comprehension mainly concentrate on brain areas related to attention and meditation. In this research, an in-depth analysis of TDL for program comprehension were analysed with additional features extracted from different areas of brain and also using subject measure (NASA TLX - National Aeronautics and Space Administration - Task Load Index) and task completion time. Three levels of task difficulty were analysed: easy, medium and difficult.

Ten students from University of Kent were asked to solve nine Java programs, which were designed with different difficulty level. The subject's cognitive load was recorded using EEG and six dimensions NASA TLX feedback obtained for each task and analysed for each subject. For EEG data three different feature methods were used for analysis: Energy, Frequency ratio and Event-related desynchronization and feed forward neural network with back-propagation training was used for classification. The results from different extracted features indicate that recorded EEG signals indeed reflect the task difficulty level for program comprehension tasks and classification network perform very well for easy and difficult tasks than easy, medium and difficult tasks. The correct classification accuracy for easy and difficult tasks was 81.38% and the accuracy significantly lowered (44.4%) for easy, medium and difficult tasks when all data was included for classification. The statistical analysis based on subject measure and task completion time also shows significant difference in task difficulty level.

Acknowledgement

Firstly, I would like to express my sincere gratitude to my supervisor Dr Palaniappan Ramasamy for his encouragement, guidance and support throughout the research. And also, I would like to thank all the participants who provided feedback for questionnaires and participated in experiments and for their precious time.

I would also like to thank my family for their immense patience and keeping me harmonious and supportive throughout the entire process.

Abbrevations

CLT	Cognitive Load Theory
WML	Working Memory Load
PCT	Program Comprehension task
BCI	Brain Computer Interface
BCM	Brian Computer Machines
EEG	Electroencephalography
EDA	Electrodermal Activity
WWS	Weighted Workload Score
GUI	Graphical User Interface
TPS	Task Presentation System
IIR	Infinite Impulse Response
WMS	Working Memory System
LTMS	Long Term Memory System
ERD	Event-Related Desynchronization
ERS	Event-Related Synchronization
TDL	Task Difficulty Level
ANN	Artificial Neural Network
FIR	Finite Impulse Response

List of Figures

List of Tables

Contents

CHAPTER 1

Introduction

Understanding how difficult a task is and how it is performed could help in many ways. For example, in computer science education, educators could be able to identify and organize learning activities of students based on their task difficulty level (TDL). According to Gerjets et al. [20], optimum learning conditions are described as providing learning conditions for learners without inducing cognitive over or under load. Of different types of memory in brain, working memory is crucial because it is related to many areas such as reasoning, learning and comprehension [26]. Thus designing learning materials according to learners cognitive load is crucial in order to achieve the optimum learning conditions. With recent popularity in Electroencephalogram (EEG) offers new and promising approaches to develop models to measure cognitive load and their applications could be seen in developing tools like adaptive learning environmentsfor students. [20].

This research is not intended to develop any such tools, instead an in-depth analysis of TDL for program comprehension tasks experienced by learners using EEG and Data Mining, which could support in developing such tools.

Motivation

There has been extensive research in EEG investigating how different features linked to various mental states and cognitive process [33] [27] [31]. But only little research carried out related to tasks difficulty for program comprehension tasks for students in computer science education. The study done by Fritz et al. assesses task difficulty level of program comprehension tasks in software development using combination of three Psycho-Physiological measures including EEG [18]. It determines the effect of task difficulty using only single channel related to attention and meditation with limited features for analysis. An in-depth analysis with more features and channels are required for better understanding as different parts of the brain contributes to different user activities. This in-depth analysis might help in developing effective learning material and tools for novice learners.

Research Objective

The aim of this research is to explore in-depth the different task difficulty level for program comprehension tasks that can be accurately measured and classified using EEG and data mining. Three levels of task difficulty level are analysed - easy, medium and difficult. This research investigates both indicators of EEG and self-reported subjective measures. This research extends the existing literature related to task difficulty and program comprehension studies performed by Fritz et.al [18], Crk et al. [11] and Klimesch [31] but not limited to procedures or methodologies used. The research investigates the following research questions,

- Is task difficulty reflected in EEG electrical signal within programming comprehension tasks?
- Which EEG features best predict task difficulty?
- Can task difficulty (specifically programming) be predicted accurately using a machine learning classifier with EEG features?

Research Contribution

The primary contribution of this research is to extend the existing research in number of ways. The event-related desynchronization and energy characteristics of different frequency bands originated from different areas of brain was evaluated for different TDL for program comprehension tasks using EEG. Features were extracted using different methods and classified using machine-learning techniques like neural networks and accuracy was measured. The results indicate that it is possible to differentiate mental task difficulty and classify using EEG and machine learning techniques for program comprehension tasks.

Research Outline

The reminder of this research is organised as follows. Background information about the research is explained in Chapter 2 and the relevant literature review is explained in Chapter 3. The experimental design, methodology and data analysis techniques employed is described in Chapter 4. The steps involved in data cleaning and transformation is explained in Chapter 5. Findings and analysis of the experiment is explained in Chapter 6. Discussion about the key findings and suggested further research are given in the Chapter 7 along with a conclusion.

CHAPTER 2

Background

This chapter details about the background information related to EEG, mental workload, feature extraction methods and classification techniques used in this research.

2.1 Cognitive load and Working memory

Any cognitive process related to task demand depends on two memory process - Working Memory System (WMS) and Long Term Memory system (LTMS) [31] and their interconnections given in Figure 2.1. Working memory is the memory that holds information temporarily and short lived and interfaces between long term memory and sensory memory [49] and responsible for cognitive process [31]. The working memory load depends on the complexity of the task user undertaken, more complex learning activities levy quite a load to the working memory and reverse for easy tasks. Throughout this research, memory workload or mental workload is defined as the amount of working memory used to process and execute a particular task.

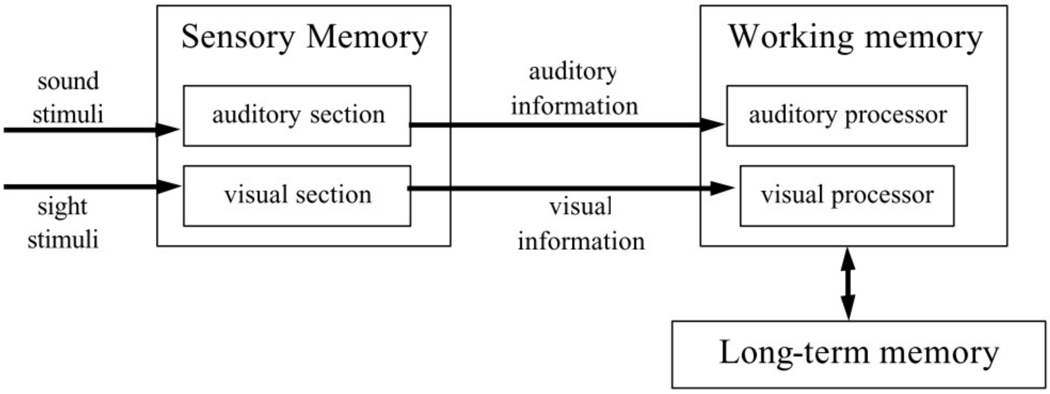


Figure 2.1: A simple human memory model [49]

2.2 Methods of Workload Detection

Currently, there are different methods to measure mental workload such as subjective, dual-task procedures and EEG etc [58] [7] [18]. Subjective and dual-task procedures likely to interrupt participants in-between the experiments and it may annoy them but produces less noisy data and provide promising results [37]. EEG is another way to measure mental workload, where the EEG device is placed on subject's head and captures brain signals based on user activities. Because of its non-obtrusive and continuous nature, it attracts researchers to adapt this method to measure the mental workload. Additional methods for measure workload could also be performed using pupil dilation and skin conductivity [18] but the results are not assuring [7]. This research uses both the subjective and EEG method to measure mental load for program comprehension tasks, which is a single task procedure.

2.3 NASA TLX-Index

NASA TLX Index, a six dimensional subjective measurement method developed by Human Resource Group at NASA for measuring cognitive load [24]. The six dimensional sub scales are mental demand, physical demand, temporal demand, performance, effort and frustration level. The workload is evaluated in two procedures for each tasks: first subject has to give their perspective sub scale rating ranging from 0–100 (divided into 20 equal intervals) and second is the sub scale weights created by forming 15 possible pair from six dimensional elements and subjects choose the most important dimension or factor contribute to the workload for performed task. The overall Weighted Workload Score (WWL) is computed from subject's rating and weight that contribute to the cognitive workload. Currently lots of research uses this method to measure cognitive load. [58] [17]

2.4 Neurophysiological Monitoring System (EEG)

EEG is a neurophysiological monitoring system used to record the electrical signals originated from brain. This monitoring system measures the current flow during transmission of information between the excited neurons. This exchange of information happens in synapse (Figure 2.2), which is present in the junction of pre-synaptic axon terminal and post–synaptic dendrite. The electrical activity generated by a group of neurons during different human activities like sleeping, task processing, relaxing etc. are captured and amplified to a level, which is sufficient for analysis [42].

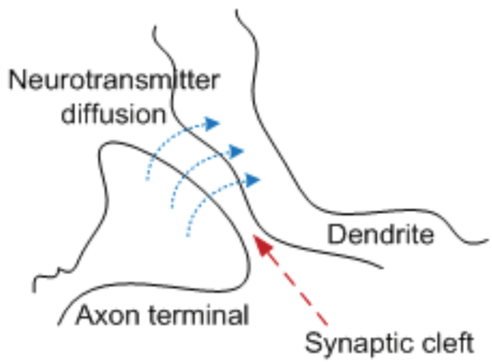


Figure 2.2: Synapses [42]

These complex oscillatory waves signal originated from different areas of brain such as frontal, parietal, temporal and occipital the brain waves can be classified in different frequency bands. According to Alotaiby et al. [1], the most useful information about the functional state of the brain is distinguished in the following frequency bands given in table 2.1.

Table 2.1: EEG frequency bands

Frequency Band	**Range**	**Detected during**
Delta Band	0-4 Hz	Deep Sleep and in babies
Theta Band	3.5-7.5 Hz	Meditation or Sleeping
Alpha Band	7.5-13 Hz	Relaxed state or when eyes closed
Beta Band	13-26 Hz	Waking or high attention state
Gamma Band	26-70 Hz	Decision making

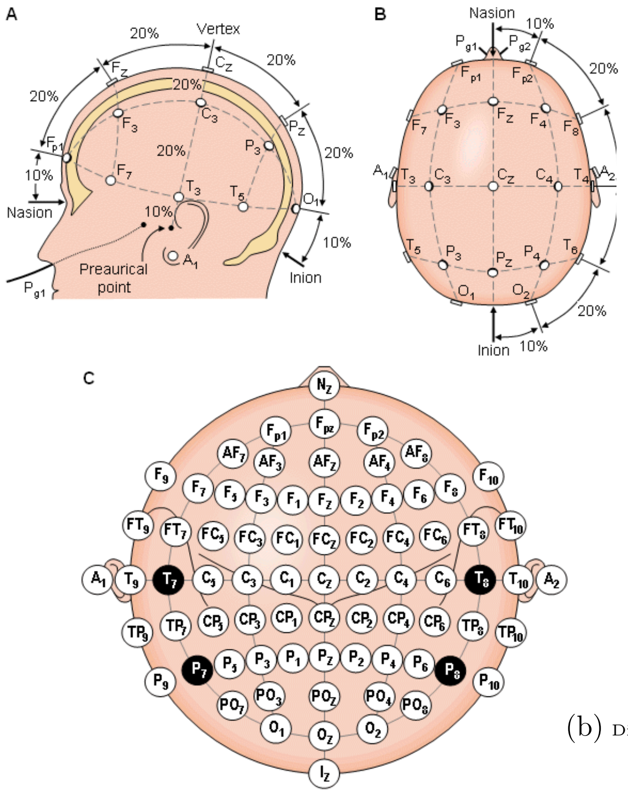
There are different ways to record signals from brain – invasive method involves surgery (placing the electrodes inside the brain using microelectrodes) that produces highly effective brain signals but is risky and the other method is non-invasive where the electrodes are placed on the surface of the head and produces weaker signals [35]. EEG system is considered to be non-invasive, and because of this non–invasive feature, the application of this system is currently increasing and it is commonly used in clinical and research field.

2.5 Placement of Electrodes (EEG)

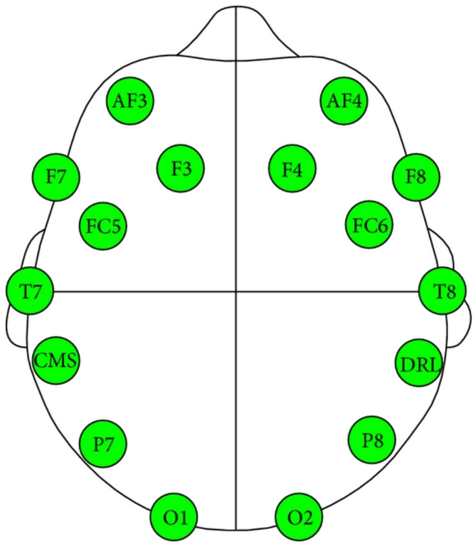
The International Federation of Societies for electroencephalography and Clinical Neurophysiology (IFSECN) recommends the 10–20 electrode system, a study method made by Dr Herbert H. Jasper for the placement of electrodes during EEG recording [28] and suggest following guidelines.

- Out of 21 electrodes, 19 electrodes are placed on specific measurement and should be proportional to the size and shape of the skull. 2 electrodes are kept as reference, one placed on ear lobes and other on mastoids.
- Electrodes should cover all parts of the head by keeping 10% or 20% distance between the edges of two electrodes. That's why the name 10 -20 electrode system.

Figure 2.3 (a) shows the placement of electrode on the head surface using 10 – 20 system. The letters A, C, F, O, P and T are abbreviated as Auricle, Central, Frontal, Occipital, Parietal, and Temporal respectively. The electrodes present in the left side of



(a) Placement of electrode on the head surface for EEG recoding using 10- 20 system. (A) Electrode placement in left side view (B) Top view electrode placement (C) Modified combinatorial nomenclature [35].



(b) Diagrammatic representation of Emotive electrode placement [56]

Figure 2.3: Electrode Placement

the head are odd numbered and the right side are even numbered. This research uses 16 channel (16 electrodes - 14 channels + 2 reference) Emotive EEG device for recording brain signals [14] and its electrode placement is given in figure 2.3 (b).

2.6 Alpha and Theta waves

The concept of Alpha and Theta frequency presented in this section is based on Klimesh study [31]. Among different frequency band originated in different parts of the brain discussed in previous section, alpha frequency band and theta frequency band play a major role in cognitive process.The author also explains that during the actual task demand, EEG oscillations in the alpha and Theta band reflects the cognitive and memory performance. This memory performance can be identified by following ways.

- Tonic increase in alpha power but a decrease in Theta power
- Tonic decrease in alpha power and but increase in Theta power during task demand (event related)

This desynchronization of alpha power is positively correlated with memory performance. Alpha frequency can be detected in the posterior and occipital region with amplitude of 50 μV (peak–peak) [55].

2.7 Individual Alpha Frequency (IAF)

The concept of Individual Alpha Frequency presented in this section is based on Klimesh study [31]. According to the author, brain maturation between subjects of different ages groups have different alpha frequency range. From childhood to teen years the alpha frequency increases and start to decrease with age. Because of this variability in age groups, part of the alpha frequency range may shift and fall outside of 8–13 Hz band. The peak alpha frequency of different age groups can be computed using below equation (Equation 2.1) [31].

$$PeakAlphaFrequency = 11.95 - 0.053 * Age \tag{2.1}$$

The adaptability of computing individual alpha frequency for experiments can be seen in many research studies. For example Crk et al. [11] study computed individual alpha frequency for each participant and adjusted the bounds of alpha band based on their age and used it as an anchor. The alpha frequency bounds are divided in 4 sub-bands defined as lower-1 alpha (L1A, ranging from IAF-4Hz to IAF-2Hz), lower-2 alpha (L2A, ranging from IAF-2Hz to IAF), upper alpha (UA, ranging from IAF to IAF + 2Hz), as well as theta (IAF - 6Hz to IAF-4Hz). By doing this, the above study suggests that it improves the functional groupings of neurons, which contributes to alpha power. This research also adapts the same functionality to adjust the bounds of alpha frequency based on age.

2.8 Digital Signal Processing

Digital signal processing is the process of manipulating the digital signal in order improve the performance and optimize the signal. It consists of different modules like filtering and noise reduction, feature extraction and classification as given in below figure 2.4. The application of signal processing can be seen in many research fields like speech processing, image processing, EEG, digital audio etc. The following sections explains the different modules of signal processing used in this research.

Figure 2.4: Steps involved in Digital Signal Processing

2.9 Digital Filtering

The concept of digital filtering techniques given in below section is based on Palaniappan [42]. According to the author, spectral digital filtering is the process of allowing components of certain desired frequencies and blocking unwanted signal components. Mostly gain of the desired frequency will be kept close to 1 and unwanted frequency to 0. The author explains that there are 4 types of digital filter in general - low-pass filter, high-pass filter, band-pass filter and band-stop filter. Attenuation and ripples are the two main characteristics of filter. Attenuation is defined as the amount by which the frequency components are attenuated and ripples are defines the amount of desired frequency deviated from gain in stop and pass band. Of many different types of digital filtering this research uses Infinite Impulse Response (IIR) filter for filtering the noisy EEG data. According to the author, IIR filters have the ability with sharper transition band than other filters like Finite Impulse Response (FIR) for the same order and its application can be seen in many researches [3].

2.10 Feature extraction methods

The following are the various feature extraction methods used to extract the features from EEG signal in this research.

Energy

According to Palaniappan [42] computing energy in different frequency bands is very often used as a features. It can be computed by using variance after setting the mean to zero. Current research on measuring cognitive load using this method produced 100 % classification accuracy in frontal channels for reading tasks.[59].

Event-Related Desynchronization (ERD)

According to Klimesch [30], ERD is a proved useful method for analysing the cognition process. The term ERD was first coined by Pfurtscheller and Aranibar [44] , where the ERD was computed by computing the percentage event related power changes between different frequency bands. ERD is also defined as extent to which neuron populations no longer oscillate in synchrony as they become activated to process [11].

Pfurtscheller and Lopes defines the steps involved in computing the ERD as, band pass filtering the EEG signal within specified frequency band and percentage band power change is computed between the relaxed state and task execution state using below equation 2.2 [31].

$$ERD = ((bandPower_{rest} - bandPower_{task}) * 100)/bandPower_{rest} \tag{2.2}$$

where, $bandPower_{rest}$ – Computed band power in relaxed state
$bandPower_{task}$ – Computed band power during task execution

Frequency Ratio

According to Fritz et al.[18], each subject has unique power distribution. Computing energy or power in different frequency bands alone for all subjects will not give enough information to differentiate the task difficulty. So the study computes frequency ratio between different frequency bands and use it as features and classified according to the TDL for program comprehension. This research also adapts the same measurements and used it as features for analysis.

2.11 Classification methods

Classification is used to classify the obtained data based on classes or groups. There are two different types of learning methods to train the classifier - supervised and unsupervised. In supervised learning, each training data is labelled with corresponding class (targets) and in unsupervised learning it is not. This research uses supervised feed forward neural networks and back propagation training for classifying TDL for program comprehension tasks.

Feed forward neural network and back-propagation training

According to [46] current classification methods depend on parametric and non-parametric multivariate analysis like discriminant and cluster analysis. The study suggests that these methods are not sufficient for non–linear data like EEG and suggest artificial neural networks for classification. The application of artificial network can be seen in many researches for classifying EEG signals related to mental load [57] [25] [34].

Artificial neural network (ANN) can be defined as an abstraction of components of human neural system. Each neuron contains some information and are interconnected. The connection between neurons controls the behaviour of the network. Feed forward neural network also known as ANN can be arranged as three or more neuron layers - one input layer, one or more hidden layers and an output layer. Usually, the hidden layer activation functions are sigmoid whereas the output layers are linear.

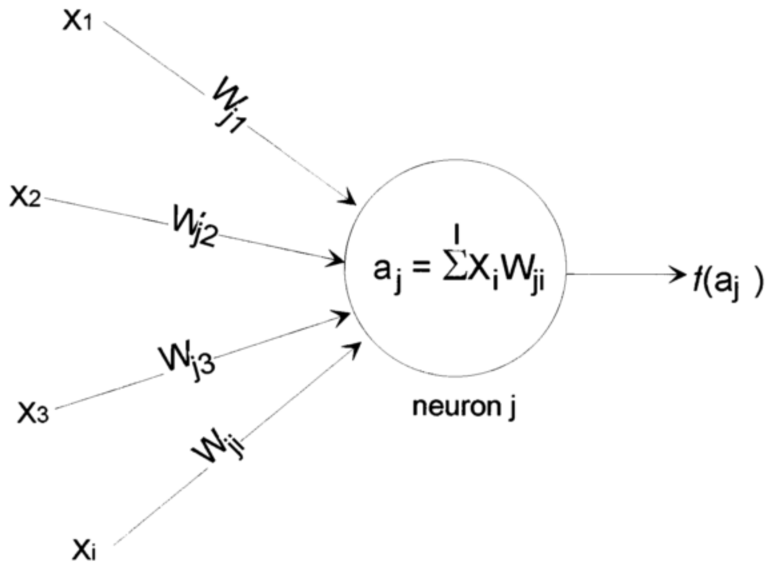


Figure 2.5: Diagrammatic representation of feed forward network [46]

Where variable X_i = Inputs to the network,

W_{ij} - Connection weights

a_j - Activation function which is defined as the summation of all inputs and connection weights and triggers output based on threshold

$f(a_j)$ - Output function.

The performance of the classifier can be computed as the percentage of correctly classified samples by the total number of samples (as given in equation 2.3).

$$\%ClassifierPerformance = (CorrectlyClassifiedSample/TotalNumberOfSamples) * 100 \tag{2.3}$$

CHAPTER 3

Literature Review

This chapter explains the current research related to program comprehension, EEG and task difficult studies.

3.1 Program comprehension and task difficulty

Program reading is defined as the process of observing computer code in some way and comprehend it [12]. Muller et al. defines program comprehension as a process where software developer understands programs by building a mental model and its relation using their domain, semantic and syntactic knowledge [38]. Program comprehension is currently an important human performance measure in software engineering and research. According to Storey [54], two kinds of program comprehension research undertaken - first is the research that involves understanding the cognitive process of how the programmers comprehend programs and the other research involves developing automated tools to improve the program comprehension.

And also two types of program comprehension have been identified. Top–down comprehension, which is first theorized by Brooks, where programmers understands the code by reconstructing the knowledge of the program in a hierarchical manner, and reconstruction starts from abstraction level hypothesis to lower level hypothesis [6]. This type of comprehension can be useful when the programmer already has knowledge about the code [53]. Another type of program comprehension is bottom–up comprehension, where programmer reads the code fully, and mentally splits into progressive layers of statements and comprehend from bottom to top level, to achieve the high level understanding [43].

The main factor that influences program comprehension are the programmer skills, program features and task characteristics. Crk et al.[11] investigates the role of expertise in program comprehension and can be measured directly. Sharif and Maletic investigates the program features such as identifier naming conventions like camel case and underscore influence program comprehension [50]. The study concluded that no difference in accuracy of identifier styles but significant improvement in time and lower visual effort.

Currently research experiments to measure program comprehension use different type of programme languages like C#, Java, python [18] [11] [23] and different methods to measure it. Some direct measurements methods include eye tracking [50], electro dermal activity [18], cerebral blood flow [40], EEG [33] and indirect measurements like interviewing and think-aloud during experiment [15], NASA TLX workload index [18] etc. This research uses program comprehension tasks written in Java using non-intrusive EEG and also subjective measurement with NASA TLX index.

For research experiments, the design of program comprehension depends on type of study and time of experiment conducted. In Fritz et al activity [18] study, the program comprehension tasks are designed to be smaller and shorter tasks (but yet limited to program comprehension and mental execution) instead of one long single tasks, as it is hard to study user activities in granular level for long tasks . The study has designed tasks to pressure subject's capability in spatial relation, visual object grouping, mathematical execution and logical program execution [51] [21] [18][11]. This same program comprehension design has been adapted in this research with above mentioned categories. This task categorical split may pressure the subject's working memory to larger amount and may lead to generate different rhythmic bands which dominate in different areas of brain.

3.2 EEG for workload measurement

Brain related measure can be defined as recording the electrical signal generated inside the brain. Non-obtrusive EEG is a easy and harmless way to capture this data. Reading of this signal can be achieved by placing the electrodes on the scalp of the participants and measure the electrical signals generated by brain activity (5–100 μV). Research shows that different frequency bands such as alpha, beta, gamma, delta and theta in the EEG data are linked to different mental states [31].

For the cognitive workload measurement, alpha and theta waves occurring in 8–13 Hz and 4–8 Hz frequency bands respectively plays a major role in cognitive performance [31]. Alpha band (8-12 Hz) is often correlated or associated of brain being inactive. This suggest that EEG signals of high alpha for low level cognitive work and low alpha for high workload [32]. The EEG frequency from 4-8 Hz (Theta band) is also associated with working memory or cognitive effort. If the task difficulty increases theta increase [29]. Research also found that delta frequency band (0-4 Hz) also correlated with working memory [59]. The study has showed a classification accuracy of 100% using time based and spectral based features taken from few frontal channel in delta frequency band.

Other studies explores that when the task solving increases, the theta frequency is reduced and the beta frequency is increased and the alpha frequency is blocked [45] [4]. Another study examined that for task difficulty manipulations theta frequency and alpha frequency are crucial [5]

Currently there are different methods to measure cognitive workload, includes direct and indirect measurement. Indirect measurement methods mostly rely on dual-task procedures [13] [7] [8] [10] or subjective methods [24], where both are obtrusive and annoy the learners during learning process. But the direct measurement which is non-obtrusive measure of working memory load is used nowadays in research [18].

In this research Emotive EEG [14] is used which is non-invasive and designed to measure the cognitive load. The captured signals will then be analysed using signal processing techniques.

3.3 EEG for Task Classification

Current researches focuses on measuring task difficulty and classify based on the difficulty level using BCI as it has the ability to translate the transition of brain mental states into states appropriate for controlling an application or devices. Two approaches has been referred in the existing literatures for accomplishing the above - operant conditioning approach and pattern recognition approach [52]. Smith explains that the pattern recognition approach uses signal processing algorithms to extract and classify the features related to the cognitive mental state and used as an input to control or communicate other devices The operant conditioning is defined as self-regulating where the user perform longer training sessions and learn how to control the brain waves to get the required output. In the pattern recognition approach, the burden of converting brain activities to digital output lies on signal processing and machine learning techniques where as in operant conditioning the burden lies on the user. This research takes the advantage of pattern recognition for processing and classifying task difficulty level as it is performed by digital techniques rather than the user.

Currently lots of research explored on developing BCI using EEG and pattern recognition approach [22]. A study made by Lee and Tan [33] uses low cost off-the-shelf EEG device to classify three cognitive tasks rest, mental, arithmetic and mental rotation tasks using machine learning algorithms and achieved mean accuracy of 84%. In another study made by Fritz et al. [18] uses single channel Neurosky (related to attention and meditation) to extract brain activities and computes different frequency band ratios and use as a features to classify the tasks difficulty for program comprehension tasks. The study achieved, for new developers, 64.99% precision and 64.58% recall and for new tasks with 83.38% precision and 69.79% recall. In a another research, Huan and Palaniappan [27] has developed a BCI model with pattern recognition approach for paralysed individuals to communicate with external environment. The study uses Keirn and Aunon's data set performed by four subjects and 5 different mental tasks using EEG and achieved 97.5% of classification accuracy for pair of tasks for each subjects.

CHAPTER 4

Experimental Design and Procedure

This research experiment is designed to measure one factor, which is the TDL based on mental workload (easy, medium and difficult) for program comprehension tasks using direct measurement (EEG) and subjective measurement (NASA TLX). Experiment was conducted and data was collected from ten postgraduate students from University of Kent, who registered for computer science courses in School of Computing. Each subject executed nine program comprehension tasks of various TDL and their respective EEG data was recorded and the subjective feedback was collected. This chapter explains the task design and procedures followed for data collection. No research or experiment is performed without issues, likewise this research also encountered some issues during the experiment and the steps undertaken are also discussed.

Ethical Approval

As this experiment involves human participation, ethical approval was obtained from the School Ethics Committee, School of Computing, University of Kent. Before starting the experiment questionnaires and forms to be used in the research was designed and submitted for ethical approval. The documents submitted for ethical approval were background questionnaire, consent form, exit questionnaire, Volunteer Information, NASA TLX form, project proposal, ethical approval application form. Only the approved forms and questionnaires were used for the experiment (Refer Appendix I from section 2 till 7).

4.1 Task Design

Experimental task was designed in a way that subjects has to understand the given program comprehension tasks and perform code execution mentally. A total of 20 questions were designed and split to 5 categories which contained five tasks of different TDL ranging from easy to difficult. In order to reduce inductive bias, the participants have to solve mentally and provide answers instead of choosing through multiple choice. All tasks were written in Java programming language. All the designed tasks could be seen in Appendix I. The program comprehension task categories are explained as follows

Spatial relation tasks

In this category, tasks were designed to test subject's spatial reasoning skills like visualizing shape objects mentally and solve tasks. This visualizing could fill the working memory with visualized objects and trigger different rhythmic bands.

Example The subject has to visualize two rectangle objects using x–axis coordinates, y–axis coordinates, width and height parameters and solve whether those two rectangles overlaps or not.

Listing 4.1: Spatial relation tasks

```
Solve the below java program
(Note: Rectangle object is defined as
        drawRectangle(x-axis, y-axis, width, height))
    class MyCanvas extends JComponent {
        public void paint(Graphics g) {
```

```
            g.drawRect (10, 10, 10, 10);
            g.drawRect(30, 50, 10, 10);
        }
    }
Will two rectangles overlap?
```

Visual object grouping tasks

This category influences the subject's working memory by remembering the swapped, mapped or sorted shape of objects group correctly and solve the task.

Example In the below program, a number of shape objects were mapped to a variable and grouped in an array in different order. The subject has to map the variable name with the shape objects correctly and output objects in order.

Listing 4.2: Visual object grouping tasks

```
    public class Main {
        public static void main(String[] args) {
            ArrayList<Object> array =
                    new ArrayList<Object> ();
            String a = "Circle";
            String b = "Triangle";
            String c = "Square";
            String d = "Triangle";
            String e= "Circle";
            String f = "Triangle";
            array.add(b);
            array.add(a);
            array.add(d);
            array.add(e);
```

```
            for (int i = 1; i < 4; i++) {
                System.out.print(array.get(i) + " ");
            }
        }
    }
What is the output of the program?
```

Mathematical execution tasks

Performing mathematical tasks is one of the best way to fill up working memory to greater amount. In this category the subject has to perform arithmetic calculations mentally and solve it.

Example In the below program, the subject has to compute the average of integers in an array.

Listing 4.3: Mathematical execution tasks

```
    public class Main {
        public static void main(String[] args) {
            int[] vars = {4, 8, 10, 12, 16, 10, 18, 2, 3, 5};
            int value = 0;
            int count = 0;

            for(int i = 0; i< vars.length; i++){
                value += vars[i];
                count++;
            }

            int temp = value/count;
            System.out.println(temp);
        }
```

```
    }
What is the output of the program?
```

Logical reasoning tasks

Performing Boolean tasks or logical reasoning tasks is also another way influence the working memory load.

Example In the below program, the subject has to solve the Boolean expressions and output the result.

Listing 4.4: Logical program execution tasks

```
public class Main {
    public static void main(String[] args) {
        int x = 3;
        int y = 5;
        int z = 8;
        x++;
        y++;
        z++;

        String result = (x > 4 && y < 6)
                ? "Dog" : (x > 6 && z > 8)? "Cat" : "Horse";

        System.out.println(result);     }
}
What si the output of the program?
```

Pre-Questionnaire

In order to eliminate the bias of researcher choosing the tasks to be used for the experiment, a pre-questionnaire consists of above designed 20 programs with feedback section were sent to 20 volunteers (Refer Appendix I). A background questionnaire was also attached along with the document to input the participant's details. The volunteers were selected outside University of Kent, who have different years of experience ranging from 6 months to 11 years in Java programming and currently working or studying (Java Experience - Mean = 30.533 months, Standard Deviation = 35.62 months). This varied years of Java experience helped in improving the effectiveness of choosing different TDL. Age range varies from 23 years and 37 years (Mean = 28.8, Standard Deviation = 4.63) and 9 were males and 6 were females.

In the pre-questionnaire document, the volunteers were instructed to solve the tasks and provide answers mentally and not to use any paper or digital devices. The volunteers also were instructed to fill the feedback section with the time-spent details and the evaluated mental rating (range 1-10, where 1 – very less mental workload used to 10 – very high mental workload used) and task difficulty level rating (range 1-10, where 1 – very easy task to 10 – impossible to solve mentally) for each task. Out of 20 volunteers, 15 of them replied back with their feedback and background details but two of the questionnaires were incomplete. Out of 20 tasks one volunteer skipped 11 tasks and another volunteer skipped 5 tasks.

Task Selection

The pre-questionnaire feedback from 15 volunteers data were gathered and consolidated based on three features, mental workload rating, task time spent (in seconds) and task difficulty level. From the consolidated result, the median value of mental workload rating, task difficulty rating and time spent for each task were determined. As the time data were in seconds, dividing each task median value with the total no of tasks normalizes it and the result is given in Table 4.1.

Based on the final rating (Table 4.1 column 6), lesser rating was selected as easy task, higher value as difficult tasks and mid value as medium task from each category. If there is a dilemma in selecting tasks because of same rating, then the tasks was selected based on researcher's knowledge on those tasks. So totally 12 questions were selected for experiment. And also from the rest of the tasks (whichever not selected), four questions were selected for practice task from each category for the subjects to familiarize about the tasks before starting the experiment.

Table 4.1: Task selection

Task category	**Question no**	**Median - Mental workload rating**	**Median Task difficulty rating**	**Time spent on tasks (Normalized)**	**Final Rating**	**Selected tasks for experiment and its difficulty type**
Mathematical mental execution tasks	1	2	2	6	10	
	2	2	2	6	10	
	3	2	2	6	10	medium
	4	1	1	4	6	easy
	5	3	3	6	12	difficult
Visual object grouping tasks	6	2	2	6	10	easy
	7	4	4	12	20	difficult
	8	4	4	12	20	
	9	4	3	12	19	
	10	3	3.5	6.5	13	medium
Logic programs	11	3	3	6	12	medium
	12	4	4	6	14	
	13	3	3	6	12	
	14	3	2	6	11	easy
	15	4	3.5	9	16.5	difficult
Spatial relation tasks	16	2	2	6	10	easy
	17	3	3	6	12	medium
	18	3	2	6	11	
	19	3	3	6	12	
	20	4	4	6	14	difficult

4.2 Subject selection

Subjects were recruited from a pool of post graduate students from School of Computing, University of Kent, who had at least 6 months of Java experience or has taken Java Programming module as a part of their current postgraduate course. Those who completed 30-45 minutes of the experiment were remunerated with £15.00 funded by University of Kent. Out of ten subjects nine were males and one was female. Subjects age ranged between 20 and 35 years (Mean = 26, Standard Deviation = 3.74).

4.3 Tools used in experiment

This research used off-the-shelf research grade Emotive Epoc head set [14] which is designed for research studies and advanced Brain Computer Interface (BCI) applications. The captured signals were processed and analysed using MATLAB R2015a. And also to present the tasks to user, a Graphical User Interface(GUI) was developed using MATLAB R2015a, which is explained in detail in next section.

4.4 Task Presentation System (TPS)

A GUI for presenting the experimental tasks to participants and capturing task responses were developed using Matlab. The below table gives an overview of the screens present in the GUI and its details. Detailed User Guide on TPS is given in Appendix F.

Table 4.2: Overview of screens present in TPS

Screen	**Details**
Welcome screen	Displays general information about the research.
Instruction screen	Displays instructions to be followed by the subjects during the experiment
Relax screen	During this screen, the subjects will relax or fill up the NASA TLX feedback for the performed tasks.There is no time limit for the subjects to relax.
Task Screen	This screen contains the task to solve and a text box for the subject to input their responses.
Thank you screen	This screen contains thank you message and indicates the end of the experiment.

This GUI not only serves as a front-end, but also communicates with the EEG device via COM port (Serial port) (connection was made using the software Virtual Serial Port Emulator [16]) and sends markers based on user activities like subjects relax state and task execution state. The below table gives the details of marker type and its value sent to EEG device during the experiment (Table 4.3).

Table 4.3: Marker type and value sent to COM port

Marker Type	Value
Start of Task	1
End of Task	2
Start of Relax state (to indicate baseline recording)	3

Subjects were demonstrated the working of Task Presentation System (GUI) and were asked to perform the practise tasks by their own using GUI in order to familiarize the tool. This helped subjects not to get stuck in between actual experiment, which may lead to physical movement and is always discouraged.Figure 4.1 gives the work flow of TPS

4.5 Instructions to participants

Before starting the actual experiment, the subjects were given instructions regarding the tools used (EEG device and Task presentation System (GUI)) and behaviour to follow during the experiment. Before experiment start, subjects were asked to sit comfortably and in relaxed state with minimal tension. As this stress might influence the data and cost the validity of data. The subjects were discouraged excessive physical movement especially blinking, excessive swallowing or any hand gestures etc. during the experiment. They were asked to focus on the presented task while solving. Subjects were ensured that the experimental procedure is harmless and have the right to stop the experiment anytime and leave.

4.6 EEG Recording

Before placing the EEG device on the subjects, the battery life of the EEG device was checked and ensured it was full. It is not safe to charge the device while in subjects head, it is best to charge it before hand and ensure it lasts for the whole experiment. Emotive Control panel (Figure 4.2), which is companion software for the wireless Emotive EEG device, displays the signal strength of each electrode and status of the battery life. In

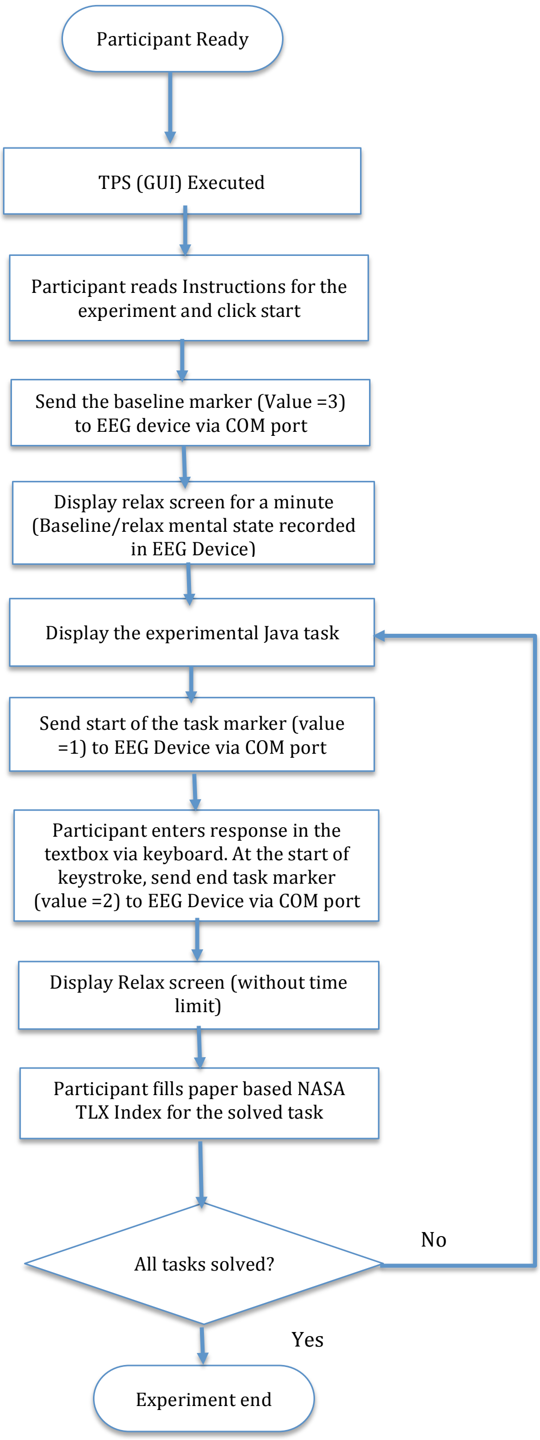


Figure 4.1: Task Presentation System work flow

order to increase the conductivity of the each channel, saline solution was added to each of the electrodes and placed it on the subject's head.

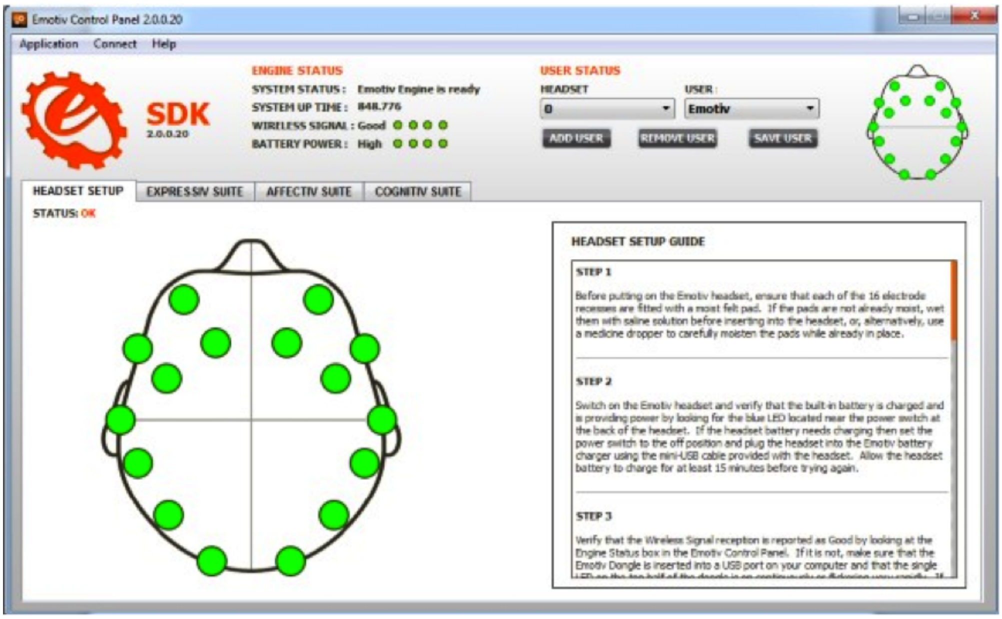


Figure 4.2: Emotive control panel indicating the signal strength of electrode and battery power

During the experiment, the signal strength was checked and adjusted till all the electrodes shows green in emotive control panel. If the signal strength is not good (corresponding electrode displays in red, orange or yellow colour) then more saline solution was added until all the electrodes showed green.

Issues Encountered

- As the emotive EEG device is wireless, the signal strength and the battery power of the device was checked in the Emotive test bench. During experiment for one subject the battery power information was shown wrongly (shown as 60% battery power), but the battery power ended up dead during the experiment and failed to record EEG data. It recorded only data for 4 tasks in total for that subject. So had to exclude the subject's data from the analysis. So next time, checking battery life was not relied on Emotive test bench software.

- For one subject the base line data was not recorded properly, so had to exclude the subject from ERD analysis but included in other analysis like Feature (energy) analysis and NASA TLX analysis.

4.7 Pilot Study

Before testing the actual research methodology, a pilot study was conducted to follow all the procedures of actual experiment. The pilot study was performed on one female volunteer who was a student from University of Kent and got 6 months of Java experience. First the subject was given a brief introduction about the session, instructions about the experimental procedure and behaviour to follow during the experiment. Next the subject was given four-practise test to familiarize with the Task Presentation System (GUI) and the tasks. Then the subject performed the actual experiment on the twelve experimental Java programs. The whole experiment lasted for one hour and ten minutes. For the pilot study, the subject was not remunerated with any money.

Issues Encountered (Pilot Study)

- This research was designed to last for no more than 30-45 minutes, but the pilot study lasted for more than an hour. So the long hours made the subject bit exhausted, even though the subject took little break between tasks.
- The signal strength was moving from ON to OFF during the experiment because of insufficient saline solution in the electrode. It was not possible to remove the EEG device during the experiment and place it in the same place as before in subject's head. This change in electrode placement may affect the validity of the result.
- The overall difference in memory workload between the easy, medium and difficult task was not much observed during the EEG data Analysis.

Solutions to the Issues (Pilot Study)

- To reduce the long hour experimental time, it's been observed that solving 12 tasks (with little breaks) made the subject bit exhausted. So it was decided to remove one category from the twelve experimental tasks. When the EEG data was analysed, the logical programs tasks doesn't much contribute to the working memory load compared to other categories, so that category was removed (3 tasks) from the experiment. The corresponding practise task was also removed. Now the experimental tasks contain only nine tasks (Refer Appendix G) and practise tasks contains only three tasks (Refer Appendix H).

- For the saline solution issue, much care was taken to keep adequate saline solution in the electrodes before the experiment that can last for the whole experimental process.

- In order to produce a greater variability between task difficulty levels, it was decided to make some easy tasks more easy and difficult tasks more difficult.

Example given in table 4.4 shows how the code was modified with more difficulty and contribute to stronger working memory load. In the below Java code falls under spatial relation category and tasks difficulty level was difficult.

Before modification, the subject has to map the variables with values and visualize the shape mentally. But now it was modified as, the subject has to multiply the x-axis and y-axis values for all the three rectangle objects with two and visualize the shape (the modified part was given in bold).This additional computation might increases cognitive workload than before.

Table 4.4: Modified code to increase task difficulty

Before code	After code

Solve the below java program and fill the output and
analysis section.
(Note: Rectangle object is defined as,drawRectangle(x-axis, y-axis, width,
height))
 	
class MyCanvas extends JComponent {
public void
paint(Graphics g) {
int x1 = 10;
int y1 = 10;
int w1 = 15;
int h1 = 15;
 	
int x2 = 15;
int y2 = 20;
int w2 = 15;
int h2 = 10;
 	
int x3 = 10;
int y3 = 45;
int w3 = 25;
int h3 = 10;
g.drawRect
(x1, y1, w1, h1);
g.drawRect(x2, y2, w2, h2);
g.drawRect(x3, y3, w3, h3);
}
}
Which two rectangles overlap?
	Solve the below java program
	(Note: Rectangle object is defined as
	drawRectangle(x-axis, y-axis, width, height))
	
	class
	MyCanvas extends
	JComponent {
	public
	void paint(Graphics g) {
	int x1 = 10;
	int y1 = 10;
	int w1 = 15;
	int h1 = 15;
	x1 = x1 * 2;
	y1 = y1 * 2;
	int x2 = 15;
	int y2 = 20;
	int w2 = 15;
	int h2 = 10;
	x2 = x2 * 2;
	y2 = y2 * 2;
	int x3 = 10;
	int y3 = 45;
	int w3 = 25;
	int h3 = 10;
	x3 = x3 * 2;
	y3 = y3 * 2;
	g.drawRect (x1, y1, w1, h1);
	g.drawRect(x2, y2, w2, h2);
	g.drawRect(x3, y3, w3, h3);
	}
	}
	Which two rectangles overlap?

4.8 Experimental Procedure

- Before the start of the experiment, the EEG device was ensured all electrodes got adequate saline solution and full battery power which can last for the whole experimental process.
- At the start of the experiment, all subjects were asked to fill out consent form and background information.
- Ensured the subjects understands the instructions and behaviour to follow during the experiment and were asked to solve three practise questions using TPS (GUI) to familiarize them with the tasks and tool.
- Once the subjects were confident, the EEG device was placed on the subjects head and adjusted till all channels receives signal properly (checked in the Emotive control panel) and subjects feel comfortable.
- 5. Once the subjects were ready, the actual experiment begins with the experimental tasks with the same TPS tool. The order of the questions presented to all participants is given in below table 4.5.

Table 4.5: Order of tasks presented to participants

Category	**Question no.**	**Task Type**
Mathematical	1	Easy
	2	Medium
	3	Difficult
Visual grouping	4	Easy
	5	Medium
	6	Difficult
Spatial reasoning	7	Easy
	8	Medium
	9	Difficult

- Baseline data was recorded during the first relax screen (before displaying first task) where subject were asked to view the relax image for one minute mandatorily.
- In between each task, a relax screen was displayed for the subject to take a break and fill the NASA TLX rating for the solved task.

- Once all the tasks were solved, the subject were asked to fill the paper based exit questionnaire, where subjects rank the task based on the difficulty level.
- Once the whole experiment was completed, the subjects were renumerated with £15.00.

4.9 Overall Issues encountered during experiment

- The main issue with the experiment is that subjects cannot retake the experiment, as they know the program tasks and its answers beforehand. Retaking the experiment affects the validity of the experimental results. So extra precautionary measures and checklists verified before starting the experiment.
- Even though subjects were instructed about not to perform excessive physical movements, but some subjects still did some physical movements like moving hands or talking aloud while solving the tasks which may have introduced noise in the EEG data.

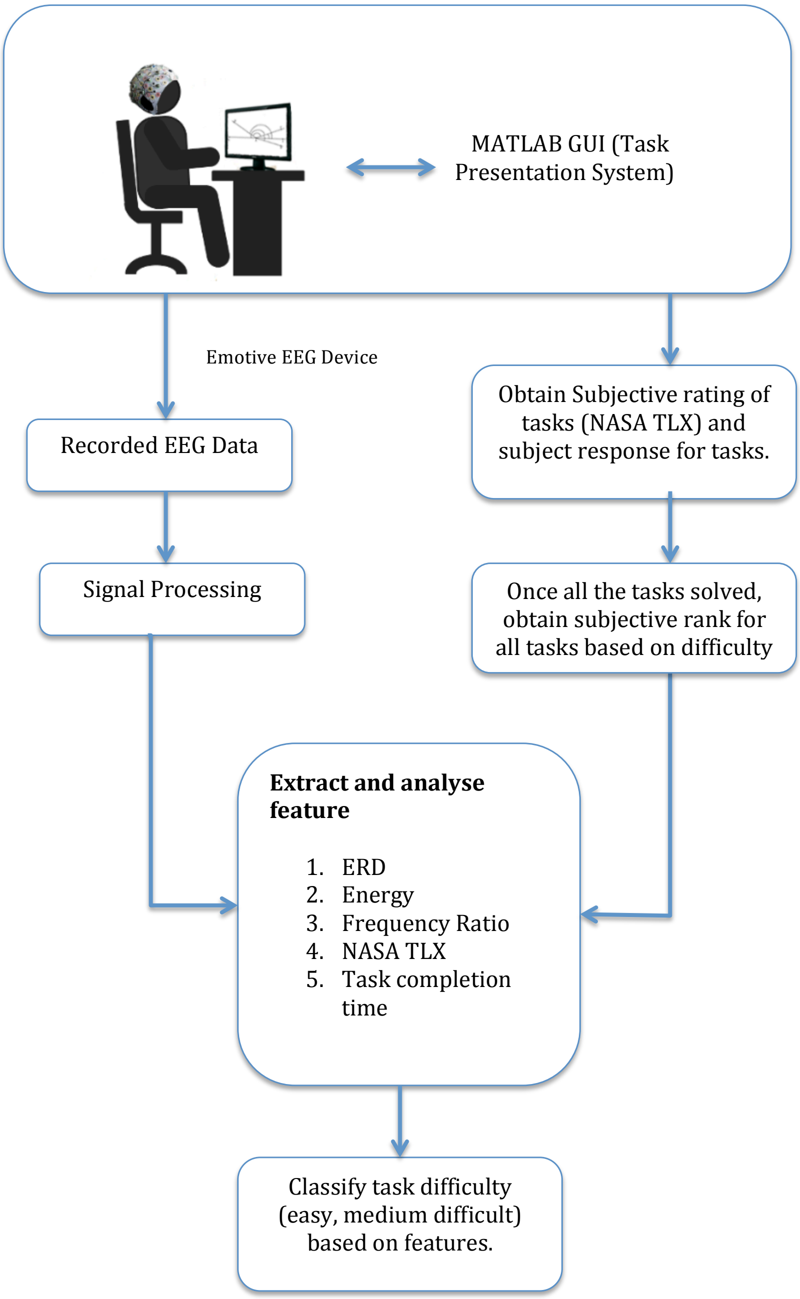


Figure 4.3: Overall experiment details

CHAPTER 5

Data Cleaning and Transformation

This chapter explains the data mining process of data cleaning and transformation of collected raw EEG data and subjective data in-order to extract useful information to support the research goal.

5.1 Extracted features for analysis

The below tables provides the list of different frequency bands (Table 5.1) and methods used for feature extraction (Table 5.2)from the raw EEG data their effects in existing literature.

Table 5.1: Frequency bands used for analysis

Frequency Bands	**Alpha** (8-12 Hz), **Beta** (12 -30 Hz), **Gamma** (30-50 Hz), **Delta** (1-4 Hz), **Theta** (4-8 Hz), **Alpha + Theta** (4 -12 Hz), **Alpha + Beta** (8-30 Hz), **All frequencies combined** (1Hz -50 Hz), **Lower-1 Alpha** (IAF-4Hz to IAF -2Hz), **Lower-2 Alpha** (IAF-2Hz to IAF), **Upper Alpha** (IAF to IAF+2 Hz), **Theta (based on IAF)** (IAF-6Hz to IAF-4Hz), Where, IAF = 11.95-0.053*Age (IAF - Individual Peak Alpha Frequency - Details in section 2.7)

Table 5.2: Feature extraction methods used for analysis

Measures	Channels computed	Measurements adapted from existing literature
Energy	Computed for all 14 channels	Biological Signal Analysis [42], Brain area related to working memory activities studies [2] [9]
Frequency band ratio	Computed for 2 channels related to attention and meditation (AF3 and AF4)	Working memory load measurement for cognitive tasks [22], Task classification using three psycho-physiological measure [18], Analysis of adaptive task using EEG [45]
ERD	Computed for all 14 channels	Expertise studies [11], Cognitive and Memory Performance studies [31], Brain area related to working memory activities studies [2] [9]

5.2 EEG Data

As raw EEG data consists of large amount of noise and invalid data, the data has to be cleaned first to remove them. The emotive raw EEG data consists of 36 columns including 14 channels data, a column to hold the marker value sent by the TPS (GUI) during the experiment and rest of the columns related to status of the EEG electrode. This research mainly concentrates on the 14 channels data (columns 3 to 16) and the marker data (column 36).

Extraction of baseline data and task data

The raw EEG data consists of a collection of baseline data (subject's relaxed state) and task data (Figure 5.1) and are divided based on the markers sent by the TPS (GUI) during the experiment (Refer table 4.3). This baseline data and task data are extracted programatically using Matlab and used for rest of the analysis.

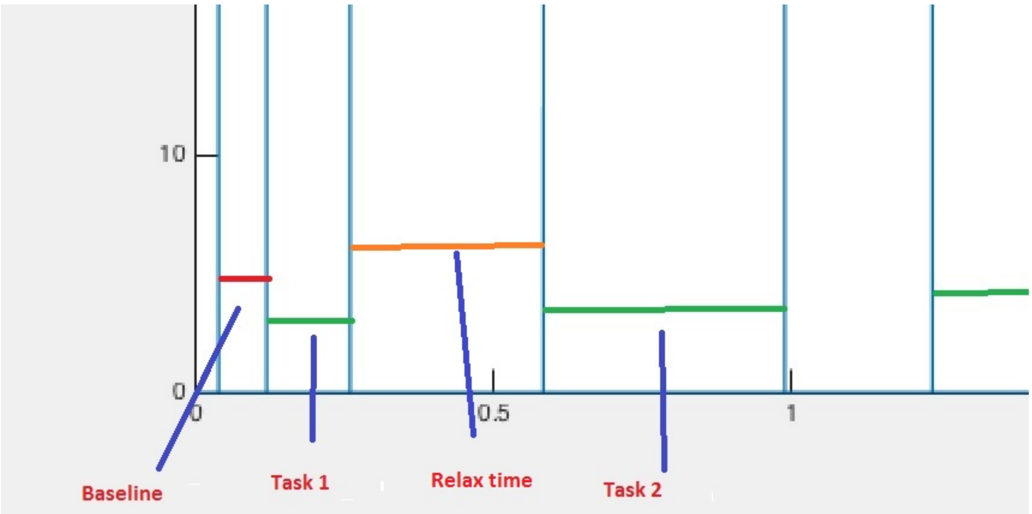


Figure 5.1: Figure shows different marker information based on participant actions like relaxing, task execution etc. during experiment.

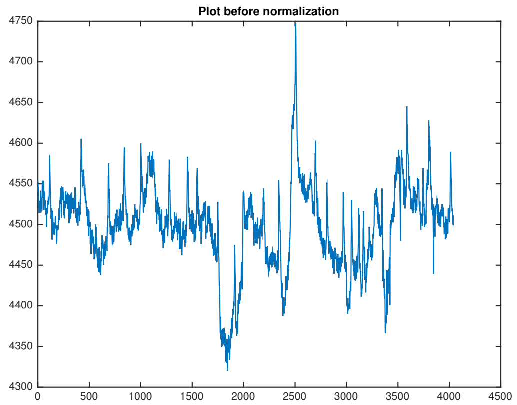
Normalization

As each subject's EEG data can differ due to several reasons, such as anatomical differences, the EEG data has to be normalized first before applying filtering techniques or extracting features. The raw EEG signal was normalized by computing mean across all channels (3 to 16 columns) and subtracted from each channel data for each subject. This method is applied for both baseline data and task data. The below sample code defines how the baseline data is normalized in this research. Figure 5.2 (a) and (b) shows the EEG signal before and after normalization.

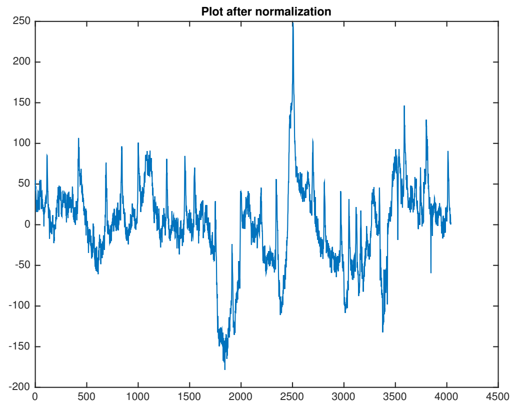
Listing 5.1: Normalization pseudocode

```
baseline_raw_data= data(3:16, baseline_marker_array(1)
                                  : start_marker_array(1));
```

```
for ch=1:no_of_channels
    baseline_data{s,ch,:} = baseline_raw_data(ch,:) -
                            mean(baseline_raw_data);
end
```



(a) Before normalization



(b) After Normalization

Figure 5.2: Normalization of EEG data

EEG Data Segmentation

As EEG data is non-stationary and non-linear, features were extracted for 1 second window and performed analysis. The whole baseline and task data were segmented to 1-second shorter window based on sampling rate of emotive EEG device (128 Hz per second). Filtering, artefact removal and feature extraction techniques were performed on those shorter segments.

Eye blink noise removal

Using the eye blink noise removal technique explained by Manoilov [36], the normalized EEG data was band-pass filtered from 0.5-3 HZ and computed the mean and maximum (peak) value for each 1 second segments. Whichever data segment contains peak that are greater than 100 times the mean of the amplitude, corresponds to eye-blink noise and was removed. Figure 5.3 shows the eye blink artefact which is represented as peak amplitude. The simple pseudo code of removal of eye-blink artefact was computed as follows

Listing 5.2: Eye blink removal pseudocode

```
FilteredData  = abs(band pass filtered 0.5-3 Hz 1 second
                                    data segment);
If Max (FilteredData  ) > (100 * mean(FilteredData  )
        Corresponds to eye-blink and remove it.
end
```

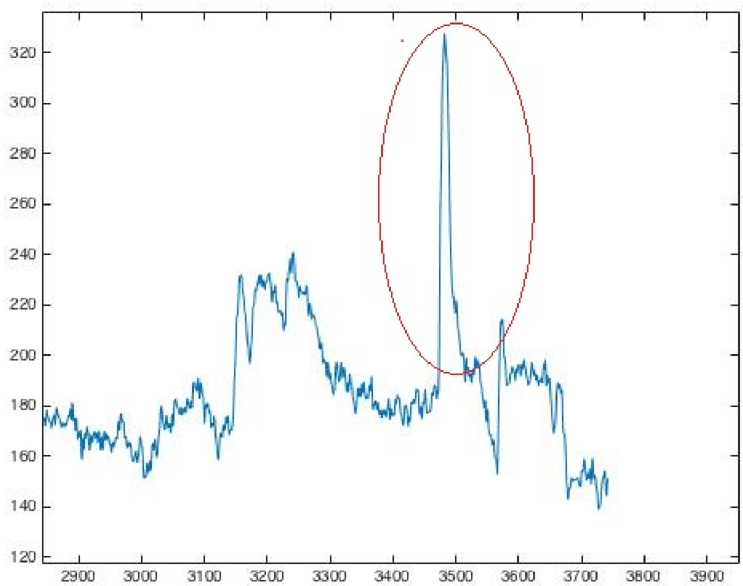


Figure 5.3: Eye blink artefact (circled in red) identified in subject 1 EEG data when performing task 1.

Filtering

Elliptic IIR filter has been used in this research to remove any unwanted noise from the raw EEG data. The filtering was done in three steps and implemented as follows (example given for 5-10 Hz signal).

- The order of the signal was obtained using Matlab function *elliporrd* with pass band of 5 to 10 Hz and stop band cut-off limit is set to 1 Hz before and after the pass band. The pass band ripples was set as 1.0 and stop band attenuation as 30 dB.

[N,Wn]=elliport([5,10]/64 ,[4,11]/64 ,1,30);
where N is the order of the signal.

- Then the order of the coefficients [B, A] was obtained using *ellip* function.

  [B, A] = ellip(N, 1, 30 , ([5,10]/64);

- In order to avoid zero-phase effect caused by non-linear distortion, where filtering is done twice in opposite direction and cancels phase effects. Matlab function *filtfilt* has been used to avoid this problem.

  FD = filtfilt(B,A, x)
  where x is the data to filter

Task grouping

For performing statistical analysis and classification the nine task were grouped into three categories based on their TDL. Based on order of tasks (Refer table 4.5) presented during the experiment, the tasks are grouped as given in the table 5.3.

Table 5.3: Grouping of tasks

Group	**Question No.**
Easy	1, 4, 7
Medium	2, 5, 8
Difficult	3, 6, 9

Energy computation

Energy was computed for each one-second window across different tasks, 14 channels, different frequency and each subject using the Matlab function *var*, then grouped based on tasks (Refer table 5.3) and performed statistical analysis (Kurskal-Wallis test) based on task difficulty level.

ERD computation

Percentage ERD computation was performed in two steps. First energy was computed (based on above method) for each one-second window across different tasks, different frequency 14 channels and each subject for baseline and trial data and % ERD was computed using the equation 2.2 given in section 2.10 (Pseudocode given below). Then values were grouped based on task difficulty and compared with statistical analysis (Kurskal-Wallis test).

Listing 5.3: ERD computation pseudocode

```
ERD_EXF(s,f,ch,q,:) = ((baseline_mean_EXF(s,f,ch,:)
-task_EXF(s,f,ch,q,t,:)) * 100) / baseline_mean_EXF(s,f,ch,:);


where s-each subject, f-each Frequency,
ch - Each channel, q - Each question, T - each time segment
```

Frequency ratio computation

For frequency ratio computation, ratio between different frequency bands were performed for the computed energy. Then values were grouped based on task and frequency band ratios and compared with statistical analysis (Kurskal-Wallis test). The below table provides the list of ratio measurements used as a feature for analysis.

Table 5.4: Frequency ratios used for analysis

Frequency ratio measurements	$\Delta(\alpha/\beta)$,$\Delta(\alpha/\gamma)$, $\Delta(\alpha/\delta)$, $\Delta(\alpha/\theta)$, $\Delta(\beta/\alpha)$, $\Delta(\beta/\gamma)$, $\Delta(\beta/\delta)$, $\Delta(\beta/\theta)$, $\Delta(\gamma/\alpha)$, $\Delta(\gamma/\beta)$, $\Delta(\gamma/\delta)$, $\Delta(\gamma/\theta)$, $\Delta(\delta/\alpha)$, $\Delta(\delta/\beta)$, $\Delta(\delta/\gamma)$, $\Delta(\delta/\theta)$, $\Delta(\theta/\alpha)$, $\Delta(\theta/\beta)$, $\Delta(\theta/\gamma)$, $\Delta(\theta/\delta)$, $\Delta(\theta/(\alpha+\beta))$, $\Delta(\beta/(\alpha+\theta))$ (Δ represents the difference to the baseline) Where, α,=Alpha, β,=Beta, γ,=Gamma, δ = Delta, θ,= Theta

5.3 Task completion time

In addition to above methods, the task completion time was also analysed. The time taken from start of the task till the subject enters first key stroke were extracted for each task and each subject. The result was then grouped based on task category and Kurskal-Wallis test was conducted and evaluated.

5.4 NASA TLX

From the collected NASA TLX feedback, the rating and the weights provided by each subjects for each tasks were extracted manually. Then the Weighted Workload Score (WWS) was computed based on the procedure given by Hart and Staveland[24]. WWL was computed for each dimension by multiplying the rating (range 0-100) and its corresponding normalized weight and summed over all dimension for a task. Then the results were grouped based on TDL. Detailed calculation which is taken from Hart and Staveland study [24] is given in Appendix I section 5. Statistical analysis (Kurskal-Wallis) was conducted on over all dimensions and also on each six dimensions.

5.5 Classification

The features extracted based on energy (12 features), frequency ratio (22 features) and ERD (12 features) were used for classification using neural networks and back-propagation training. After extracting all features for 1-second window for all tasks and all subjects and grouped based on TDL, produced 1493 samples for easy group, 2673 samples for medium group and 3236 samples for difficult group. Feed forward network was created using Matlab function *feedforwardnet* and trained using Levenberg-Marquardt back propagation algorithm with 100 epochs with 0.01 learning rate. The total records were split into 10 folds and cross-validated. 75% of the records are used for training, 15% for validation and 15% for testing. The validation records were used to avoid network over-fitting. The performance of the network was evaluated using Mean Square Error function which is the averaged square error between the target outputs and predicted outputs.Using the input and target data, correct classification accuracy was computed using the equation 2.3 given in section 2.11 for each fold and mean accuracy was computed.

CHAPTER 6

Result and Analysis

The experiment data was analysed and the data results were presented in this chapter. The first part of the chapter is concerned with the results obtained from EEG data and the second part deals with the results obtained from the subjective measure NASA TLX and task completion time.The last part is presented with classification results.

6.1 EEG Data

Energy

Kurskal-Wallis test was conducted for features extracted based on energy method to find any difference between TDL for each subject and the detailed plots and results can be seen in Appendix A and B. From the results, its been observed that there was a significant difference between TDL (easy, medium and difficult) for each subject and all features ($P < 0.01$) (Refer Appendix B) which indicates that TDL is reflected in EEG signals.

Since Kurskal-Wallis test shows significant difference in all three tasks, further exploration was done to find if there is any pattern, trend observed in different task categories for each feature. Out of 8 only 5 subjects showed a pattern of having easy task value less than medium task in alpha band (8-12 Hz) (average mean energy of 5 subjects - easy-298.31, medium-2930.88, $p<0.01$) and 6 subjects shows pattern of easy < medium in alpha+theta band (4-12 Hz) (average mean energy of 6 subjects easy-2368, medium-18760, $p<0.01$). Additionally, 5 subjects show a significant pattern of easy tasks having lesser mean value than difficult tasks (easy <difficult) in delta band (12-30 Hz) (average mean energy of 5 subjects - easy - 448.06, difficult-822.49, $p<0.01$).

As one of the objective of this research is to find if there is any significant difference and pattern in task difficulty for overall subjects and each feature. The statistical results (Kurskal-Wallis test) shows that there was significant difference between different TDL for all subjects and every feature ($p<0.01$). Significant pattern was observed for easy and medium tasks across all features and also observed that overall energy for medium task was significantly high for all features than other tasks as shown in figure and table 6.1 and figure 6.1

Table 6.1: Over all mean energy and statistical analysis

Feature name	Mean Energy			P Value	Chi-Squ	Degree of freedom	No. of samples
	Easy	Medium	Difficult				
Lower Alpha 1	694.37	3112.15	283.23	2.89E-42	191.29	2	(1493, 2672, 3235)
Lower Alpha 2	316.09	2061.79	239.31	1.02E-32	147.33	2	(1493, 2672, 3235)
Upper Alpha	415.74	1201.03	183.09	1.16E-38	174.70	2	(1493, 2672, 3235)
Theta (IAF)	368.96	1875.91	239.22	4.53E-30	135.13	2	(1493, 2672, 3235)
Delta	4346.02	8086.54	1028.12	8.07E-12	51.08	2	(1493, 2672, 3235)
Alpha	984.72	3268.07	522.90	1.11E-32	147.14	2	(1493, 2672, 3235)
Beta	1708.58	10935.1	1687.86	7.50E-37	166.36	2	(1493, 2672, 3235)
Theta	1156.40	6413.18	734.85	7.29E-23	101.94	2	(1493, 2672, 3235)
Gamma	603.69	4173.61	1201.92	5.71E-79	360.32	2	(1493, 2672, 3235)
Alpha + Theta	2355.24	11072.41	1422.39	3.29E-19	85.11	2	(1493, 2672, 3235)
Alpha + Beta	2804.70	14503.13	2358.37	2.39E-24	108.78	2	(1493, 2672, 3235)
All frequencies	11192.8	33338.66	5475.67	7.79E-20	87.99	2	(1493, 2672, 3235)

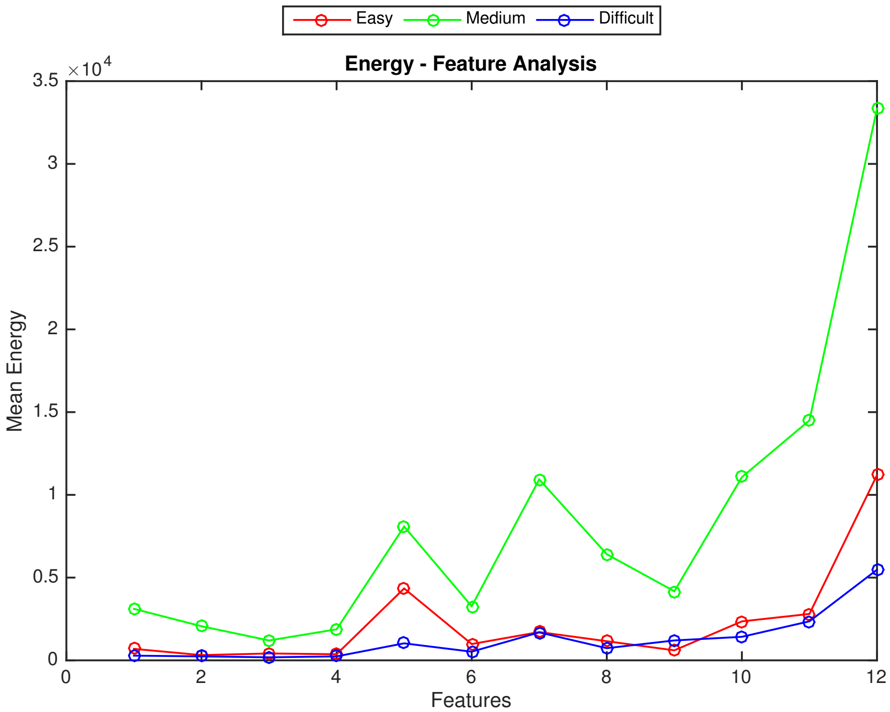


Figure 6.1: Overall Energy feature analysis. X-axis features are represented as (1-Lower 1 Alpha, 2-Lower 2 Alpha, 3-Upper Alpha, 4-Theta (IAF), 5- Delta, 6-Alpha, 7-Beta, 8-Theta, 9-Gamma, 10-Alpha + Theta, 11-Alpha + Beta, 12-All frequencies)

ERD

According to Klimesh [31], the desynchronization related to events (ERD) are represented as positive values and synchronization (ERS) is reflected as negative values. The overall results given in table 6.2 shows that nearly all feature values are negative than positive indicates synchronization than desynchronization. No pattern or trend was observed between easy, medium and hard for overall subjects. (Please refer table 6.2 and figure 6.2).But when the results are observed for each subject, some subjects shows positive ERD values (example Subject ID 8 and 10).(Results and plots given in Appendix A and C).

Patterns, trends were observed for each feature across different subjects where, in upper alpha band 2 subjects shows a significant pattern of easy $<$ medium $<$ difficult (Average over 2 subjects, easy - 66.89, medium - 86.06, difficult - 92.18). Additionally in Lower 1 Alpha and Lower 2 Alpha band there was a significant pattern of easy task having lesser mean ERD value than difficult task (easy $<$ difficult) for 3 subjects was also observed.(Lower 1 Alpha Average over 3 subjects, easy - 76.52, medium - 76.49, difficult - 83.13, Lower 2 Alpha average over 3 subjects, easy - 53.92, medium - -641.38, difficult - 61.27)

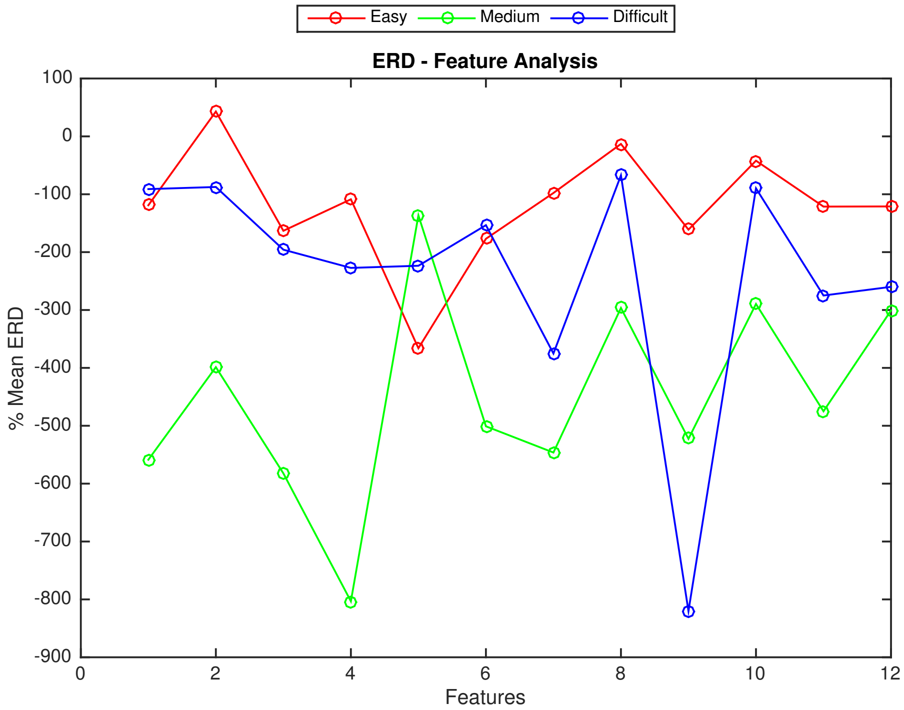


Figure 6.2: Overall ERD feature analysis X-axis features are represented as (1-Lower 1 Alpha, 2-Lower 2 Alpha, 3-Upper Alpha, 4-Theta (IAF), 5-Delta, 6-Alpha, 7-Beta, 8-Theta, 9-Gamma, 10-Alpha + Theta, 11-Alpha + Beta, 12-All frequencies)

Table 6.2: Overall ERD and statistical analysis

Feature name	Percentage Mean ERD			P Value	Chi-Squ	Degree of freedom	No. of samples
	Easy	Medium	Difficult				
Lower Alpha 1	-119.05	-558.69	-91.09	3.31E-16	71.28	2	(1493, 2672, 3235)
Lower Alpha 2	42.95	-398.71	-87.43	2.28E-08	35.19	2	(1493, 2672, 3235)
Upper Alpha	-163.14	-582.22	-195.75	1.07E-18	82.75	2	(1493, 2672, 3235)
Theta (IAF)	-108.71	-803.23	-227.24	1.23E-16	73.27	2	(1493, 2672, 3235)
Delta	-366.04	-136.48	-223.5446	1.45E-28	128.19	2	(1493, 2672, 3235)
Alpha	-176.94	-500.91	-153.4092	1.22E-24	110.12	2	(1493, 2672, 3235)
Beta	-97.81	-546.26	-375.1252	6.19E-14	60.82	2	(1493, 2672, 3235)
Theta	-13.08	-296.44	-67.33	1.96E-08	35.49	2	(1493, 2672, 3235)
Gamma	-161.15	-522.57	-820.14	7.82E-27	120.22	2	(1493, 2672, 3235)
Alpha + Theta	-42.04	-289.54	-89.77	1.59E-16	72.75	2	(1493, 2672, 3235)
Alpha + Beta	-121.32	-474.97	-274.94	3.07E-13	57.62	2	(1493, 2672, 3235)
All frequencies	-121.05	-300.78	-259.97	2.90E-08	34.71	2	(1493, 2672, 3235)

Frequency ratio

The frequency ratios between different frequency bands were computed for all subjects and two channels related to attention and meditation (AF3 and AF4) and performed statistical analysis (Kurskal-Wallis) for different TDL. The overall and each subject results shows significant difference between different TDL ($p<0.01$), which indicate different TDL is reflected in EEG data. For plots and tables for each subject given in Appendix A and D). The over all results is given in table 6.3 and figure 6.3.

When compared between each subjects (refer Appendix D), 4 subjects shows a pattern of having ratio values lesser for easy tasks followed by medium and difficult tasks in alpha and gamma band ratio ($\Delta(\alpha/\gamma)$) (Average over 4 subjects easy - 0.97, medium - 1.81, difficult - 2.22). And also all 8 subjects shows a pattern between easy and medium tasks, where easy value lesser than medium in theta and alpha + beta band ratio ($\Delta(\theta/(\gamma)$) (average over 8 subjects easy - 1.05, medium - 1.61, difficult - 1.60)and in theta and delta band ratio ($\Delta(\theta/\delta)$) (Average over 8 subjects easy - 4.04, medium 5.61, difficult - 5.6).

No pattern was observed for all three task categories for overall results as above method. Please refer below table 6.3.

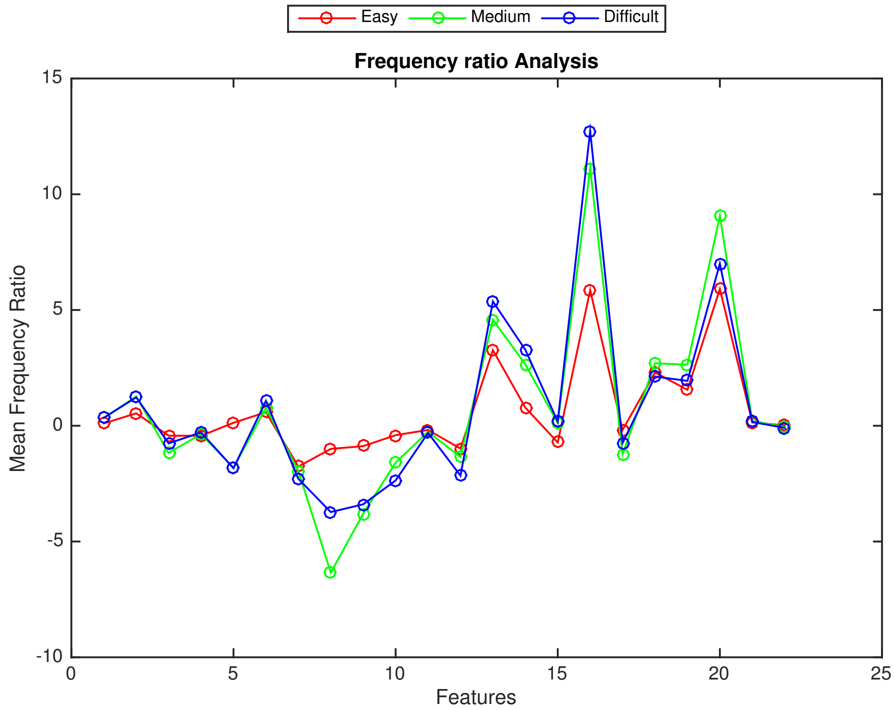


Figure 6.3: Overall frequency ratio feature analysis. X-axis features are represented as (1-$\Delta(\alpha/\beta), 2 - \Delta(\alpha/\gamma), 3 - \Delta(\alpha/\delta), 4\delta(\alpha/\theta), 5\Delta(\beta/\alpha), 6\Delta(\beta/\gamma), 7(\beta/\delta), 8\Delta(\beta/\theta), 9\Delta(\gamma/\alpha), 10\Delta(\gamma/\beta),, 11, \Delta(\gamma/\delta), 12\Delta(\gamma/\theta), 13-\Delta(\delta/\alpha), 14-\Delta(\delta/\beta), 15-\Delta(\delta/\gamma), 16-\Delta(\delta/\theta), 17-\Delta(\theta/\alpha), 18-\Delta(\theta/\beta), 19-\Delta(\theta/\gamma), 20-\Delta(\theta/\delta), 21 - \Delta(\theta/(\alpha+\beta)), 22 - \Delta(\beta/(\alpha+\theta))$ Where, Δ represents the difference to the baseline, $\alpha = Alpha, \beta = Beta, \gamma = Gamma, \delta = Delta$.

Table 6.3: Overall Mean frequency ratio and statistical analysis

Feature name	**Mean Frequency Ratio**			**P Value**	**Chi-Squ**	**Degree of freedom**	**No. of samples**
	Easy	Medium	Difficult				
$\Delta(\alpha/\beta)$	0.10	0.33	0.34	7.17E-44	198.68	2	(1493, 2672, 3235)
$\Delta(\alpha/\gamma)$	0.53	1.26	1.22	7.42E-65	295.32	2	(1493, 2672, 3235)
$\Delta(\alpha/\delta)$	-0.42	-1.17	-0.76	5.17E-73	332.89	2	(1493, 2672, 3235)
$\Delta(\alpha/\theta)$	-0.43	-0.35	-0.28	9.05E-06	23.22	2	(1493, 2672, 3235)
$\Delta(\beta/\alpha)$	0.12	-1.81	-1.83	5.30E-68	309.81	2	(1493, 2672, 3235)
$\Delta(\beta/\gamma)$	0.58	0.76	1.07	5.63E-20	88.64	2	(1493, 2672, 3235)
$\Delta(\beta/\delta)$	-1.76	-1.95	-2.29	1.36E-93	427.66	2	(1493, 2672, 3235)
$\Delta(\beta/\theta)$	-0.99	-6.34	-3.72	5.47E-129	590.66	2	(1493, 2672, 3235)
$\Delta(\gamma/\alpha)$	-0.88	-3.79	-3.39	3.90E-109	499.24	2	(1493, 2672, 3235)
$\Delta(\gamma/\beta)$	-0.40	-1.59	-2.36	2.75E-103	472.30	2	(1493, 2672, 3235)
$\Delta(\gamma/\delta)$	-0.18	-0.25	-0.30	1.87E-35	159.92	2	(1493, 2672, 3235)
$\Delta(\gamma/\theta)$	-0.98	-1.34	-2.15	1.27E-75	344.90	2	(1493, 2672, 3235)
$\Delta(\delta/\alpha)$	3.26	4.58	5.36	5.56E-86	392.61	2	(1493, 2672, 3235)
$\Delta(\delta/\beta)$	0.78	2.65	3.28	9.51E-78	354.69	2	(1493, 2672, 3235)
$\Delta(\delta/\gamma)$	-0.71	0.12	0.17	9.82E-47	211.87	2	(1493, 2672, 3235)
$\Delta(\delta/\theta)$	5.85	11.12	12.73	4.48E-54	245.68	2	(1493, 2672, 3235)
$\Delta(\theta/\alpha)$	-0.17	-1.23	-0.73	2.76E-61	278.88	2	(1493, 2672, 3235)
$\Delta(\theta/\beta)$	2.30	2.6987	2.1266	1.42E-11	49.9509	2	(1493, 2672, 3235)
$\Delta(\theta/\gamma)$	1.5519	2.63	1.93	2.62E-60	274.38	2	(1493, 2672, 3235)
$\Delta(\theta/\delta)$	5.90	9.06	7.01	2.09E-30	136.67	2	(1493, 2672, 3235)
$\Delta(\theta/(\alpha+\beta))$	0.14	0.21	0.16	4.66E-31	139.68	2	(1493, 2672, 3235)
$\Delta(\beta/(\alpha+\theta))$	0.02	-0.06	-0.09	8.55E-36	161.49	2	(1493, 2672, 3235)

6.2 Task completion time

Overall mean time (given in table 6.4 and figure 6.4) and mean time took by each subject (given in table 6.5) to complete the task was analysed for different task categories. As expected both the results indicate that the easy tasks took lesser time to complete followed by the medium and difficult tasks. The Kurskal-wallis test was also conducted on overall completion time to evaluate the difference among the different task difficulty levels. The Kurskal-wallis H test showed that there was a statistically significant difference between easy task, medium task and difficulty task with $\chi^2(2, N = 24) = 17.69$, p = 0.00014445 ($p < 0.05$).

Table 6.4: Over all mean time and mean rank (statistical analysis) of different task difficulty level

Mean Rank			Mean time (in seconds)			P Value
Easy	Medium	Difficult	Easy	Medium	Difficult	0.00014445
22.5	39.71	47.29	62.65	111.80	135.23	

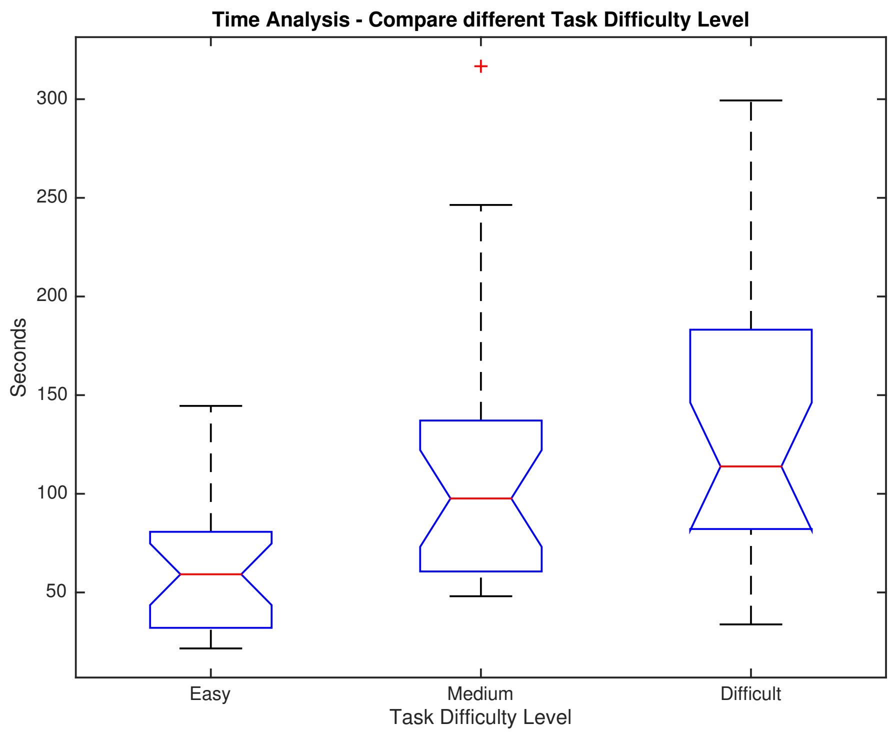


Figure 6.4: Box plot for overall statistical time analysis of different task difficulty level

Table 6.5: Per subject mean time (in seconds) of different task difficulty level

Subject ID	**Mean Time (in Seconds)**		
	Easy	**Medium**	**Difficult**
1	85.36	99.39	113.01
2	35.42	68.67	87.93
3	92.73	122.32	180.07
4	38.46	75.13	97.46
5	35.77	79.18	210.12
6	31.50	62.22	75.45
7	93.59	153.94	174.22
8	88.33	233.48	143.58

6.3 Subjective measure results (NASA TLX)

Table 6.6 and figure 6.5 shows the overall WWS computed from subjects perceived task difficulty (NASA TLX) for different task. The overall Kurskal-wallis H test result (all six dimensions combined) shows there is significant difference between TDL with $\chi^2(2, N = 30) = 18.8813$, p = 0.000079427 (p <0.01).

Kurskal-Wallis test was conducted for each dimension or sub scale as well and result shows that there was significant difference dimension mental demand, temporal demand, frustration and effort (p <0.01) for different TDL. Mental demand shows significant discrimination for easy and other tasks than all three categories. Please see table 6.7 and figure 6.6.

Table 6.6: Overall NASA TLX index mean weighted workload for different task difficult level.

Mean Rank			**Mean Weighted Workload**			**P Value**
Easy	Medium	Difficult	Easy	Medium	Difficult	7.94e-05
26.41	48.59	49.99	28.6	53.2	54.7	

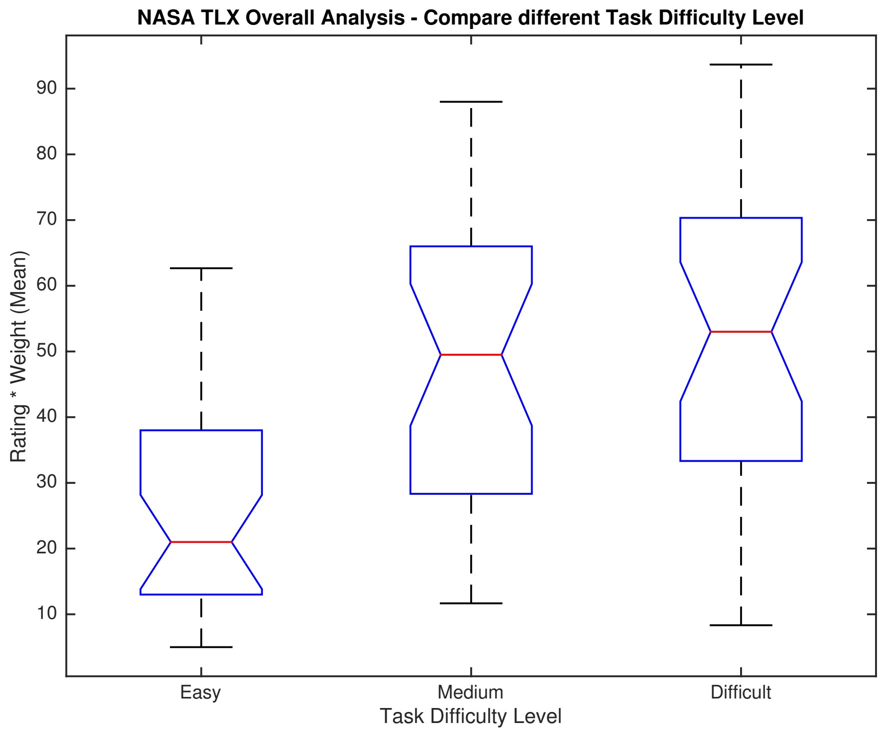


Figure 6.5: Boxplot of overall NASA TLX index mean weighted workload for different task difficult level

Table 6.7: NASA TLX - Feature based analysis

Sub scale	**Mean**			**Mean Rank**			**P Value**
	Easy	Medium	Difficult	Easy	Medium	Difficult	
Mental Demand	7.5222	15.8333	15.3556	28.4833	53.7833	54.2333	7.0757e-05
Physical Demand	0.14444	0.15556	0.3	48.15	42.4833	45.8667	0.59114
Temporal Demand	5.8222	10.2444	10.4111	32.5	52.75	51.25	0.0036795
Performance	4.8444	5.4667	6.2333	40.2333	46.2667	50	0.34324
Effort	7.2778	12.6333	13	33	50.8167	52.6833	0.0055345
Frustration	0.8	4.2556	4.6889	33.6667	53.1833	49.65	0.0072855

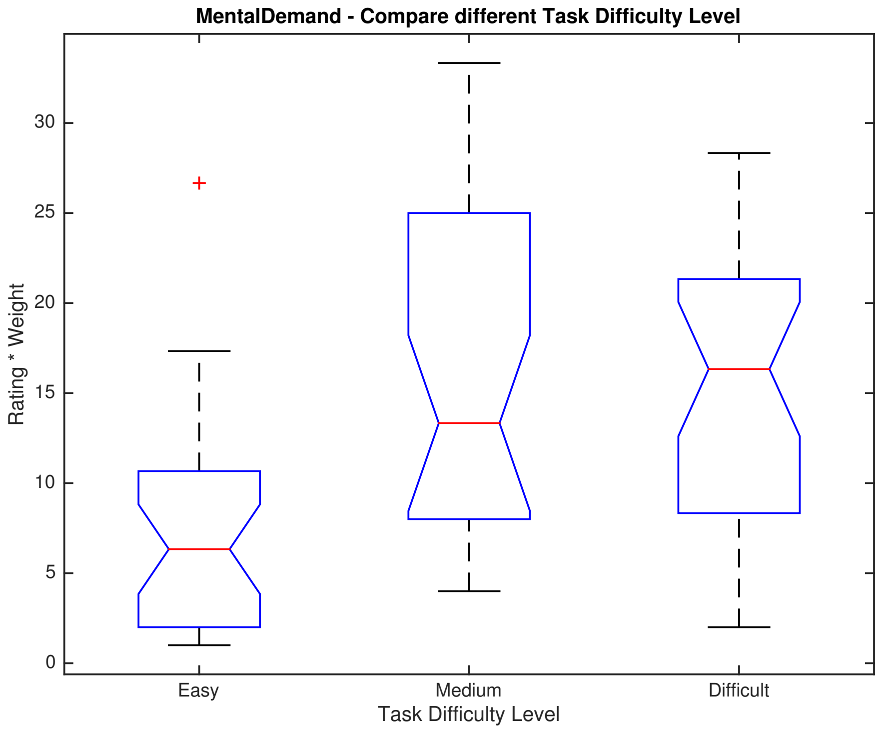


Figure 6.6: Boxplot of mental demand NASA TLX index mean weighted workload for different task difficult level

6.4 Classification

Using the all the features extracted based on energy, ERD and frequency ratio (easy group - 1493 samples, medium group - 2673 samples and difficult group - 3236 samples), classification was performed by following methods based on how data was split into 10 folds.

Method 1: Dataset was cropped into n number of samples (based on task which got lesser samples for a subject) extracted from each task group and from each subject and converted to 10 folds, so that all folds have equal number of samples from all subjects and all task group. The correct classification accuracy was computed for all three classes(easy, medium and difficult) and for easy and difficult class (for results refer table 6.8 column 1 and 2). The accuracy for all three-task categories produced 34.63% while for easy and difficult task the accuracy improved by 52.5%.

Method 2: All data from easy, medium and difficult task were used and split into 10 folds and performed classification accuracy. For all three classes accuracy was 44.4% and for easy and difficult categories the accuracy very well improved by 81.38%. (for results refer table 6.8 column 3 and 4)

The results from two methods indicate that the classification network perform well for classifying easy and difficult tasks than all three task categories and only when all information from all subjects samples were included for classification. The regression and performance plots are given in Appendix E.

Table 6.8: Classification accuracy result for 10 folds and mean accuracy based on two methods

Fold	**Method 1 (Easy vs Medium vs Difficult) - 2160 samples**	**Method 1 (Easy vs Difficult) - 1440 samples**	**Method 2 (Easy vs Medium vs Difficult) - 4479 samples**	**Method 2 (Easy vs Difficult) - 2985 samples**
1	31.94	53.47	38.10	85.16
2	33.79	56.25	41.89	76.10
3	37.03	58.33	56.08	86.04
4	37.5	56.25	35.81	74.84
5	35.18	41.66	33.10	95.13
6	34.25	49.30	49.72	83.08
7	40.27	52.77	46.08	76.95
8	31.94	58.33	54.32	69.55
9	33.33	50	37.29	96.40
10	31.01	48.61	51.62	70.55
Mean	**34.63**	**52.5**	**44.40**	**81.38**

CHAPTER 7

Discussion

On looking at the results, the overall statistical analysis shows a significant difference for all EEG features ($p<0.01$) indicates the TDL is indeed reflected in EEG signal. And also patterns or trends observed from 3 feature extraction methods, energy, frequency ratio and ERD shows significant trend for easy and medium tasks observed for energy and not for other methods. But a speculation (explained in limitations section 7) identified that physical movement artefact might have contributed to this higher energy level in medium tasks. Further analysis can be focused on removing these artefacts for better pattern detection.

When compared for each subject, computed mean energy shows significant pattern for 6 subjects for easy and medium tasks (easy$<$medium) in alpha and alpha + theta band and also pattern identified for easy and difficult tasks (easy $<$ difficult) in delta band for 5 subjects. Similar result was observed in the research conducted by Zarjam et al. [59] mentioned that the mean energy of EEG signal for delta band was observed to reflect task difficulty of reading task (low, medium and high task difficulty). In this research data was analysed only for 5 participants. In this study, the energy level decreased when there is an increase in working memory load. These stages were observed in frontal channels ((Fp1, Fp2, AF3, AF4, Fz, F3, and F4) and classification using SVM (Support Vector Machines) provides accuracy of 97-100% for computed energy. In this research, the mean energy was computed for all 14 channels and classification performance provided lesser accuracy than the study. Further analysis can be directed in focusing only on frontal channels to increase the accuracy.

Computed Mean % ERD shows significant pattern for 3 participants in upper alpha (pattern easy $<$ medium $<$ difficult), lower 1 Alpha and lower 2 alpha band (pattern easy $<$ difficult) when compared for TDL for each subject. Similar result was observed in program comprehension expertise studies done by Crk et. al. [11], where the author observed a significant difference in Upper Alpha, lower -1 alpha and lower 2 alpha band for independent variables class level based on java experience and correctness.

Computed mean frequency ratio shows pattern for different TDL in (alpha/gamma) ratio (4 subjects) for all three tasks and (theta /gamma) ratio for easy and medium tasks (8 subjects) when compared between individual subjects. Even in the available existing literature for task difficulty [18], no research looked into patterns for different frequency bands, however they performed classification based on the frequency ratio measurement.

Both NASA TLX and task completion time showed exact difference between TDL for overall and also individual subjects ($p<0.01$). In existing literature NASA TLX has been used as a non-physiological measure to discriminate different cognitive load for different programming language [58]. Based on the result we can infer that the task we have created has a good TDL discriminatory capacity. Because of the time constraint we were not able to conduct some correlation statistics between the time and EEG features which was already performed in expertise studies [11]. This might throw some insight on task completion time oriented program comprehension task design.

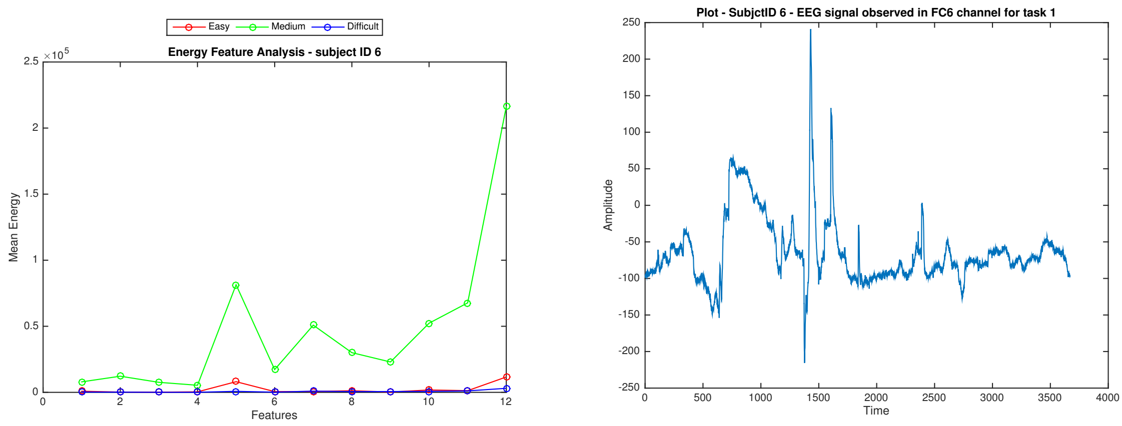
In this research neural network was selected as a choice of classifier. Initially all the data were included for classification and analysed. The result showed 44.40% of accuracy for all three TDL classes and 81.38% accuracy for easy and difficult classes. During this classification it was noticed that there was an uneven samples for each task difficulty category (easy- 1493 samples, medium - 2673 samples, difficult - 3236 samples). Because of this, the training of network might always lean towards the dominant class. So it was decided to cropped dataset (based on tasks which got lesser samples) into n number of samples from each task difficulty class and task category. So the number of samples for classification reduced significantly (Refer table 6.8). Now the classification accuracy was significantly decreased to 34.63% for all three task classes and 52.5% for easy and difficult classes. One possible reason might be that important information might have lost while cropping the data to n number of samples. In further studies techniques like duplicating minority class examples, snowball technique and multi dimensional Gaussian model of data noise can be applied[39] as suggested in the study.

In other research different classification techniques like naive bias, SVM, Bayesian network, ANN with 6th order AR coefficients with Burg's algorithm [33] [18] [59] [27] could be performed to improve accuracy. Because of the time constraint with regards to this masters dissertation further detailed classification analysis can be undertaken in the future.

Limitations of research

This research was limited by noise and experiment design procedure. Firstly, the feedback for pre-questionnaire could be obtained from student rather than from software professionals. This could have given exact discrimination of TDL for program comprehension according to student level. Secondly, from the pre-questionnaire feedback, there was some discrimination in the TDL observed, but discrimination level was not that significant (Refer table 4.1).

From the result, it can be observed that no pattern or trend found for overall subjects for different TDL. One speculation for this random patterns might be due excessive physical movement like nodding, rolling pen in their hand or think-aloud while solving task, which was observed during the experiment and EEG data (See figure 7.1a and 7.1a). Another speculation might be, EEG data doesn't provide information about whether the subject is guessing answer or actually solving it. These speculations might have cost validity of result.



(a) (b)

Figure 7.1: (a)Figure shows high level of energy for medium task for all features for Subject ID 5, might have physical movement artefacts. X-axis features are represented as (1-Lower Alpha 1, 2-Lower Alpha 2, 3-Upper Alpha, 4-Theta(IAF), 5-Delta, 6-Alpha, 7-Beta, 8-Theta, 9-Gamma, 10-Alpha + Theta, 11-Alpha + Beta, 12-All frequencies combined)
(b) EEG signal observed in FC6 channel for subject ID 5 shows noise related to physical movement.

The EEG data was performed as one time assessment. Familiarization of the research was given few minutes before the actual experiment and this might have caused noise and non-adherence to proper instruction, might compromise the quality of the data. A familiarization session followed by recording in a different day might improve the quality of the data. However this concern is not expressed in any other EEG research.

Reflection on learning

Being new to research, Cognitive Neuroscience, Signal processing, Machine learning and Matlab programming, the student researcher found that the challenges it provided brought the best out of her. The reason behind choosing this research is that, this provided the opportunity to learn entire research's life cycle from experiment design, literature review, implementation, data collection, analysis and preparing report. It was very intriguing and exciting experience and provided the researcher with a steep learning curve. This research project also helped in improving the researcher's organization skills and time management. The researcher learnt new programming language MATLAB and had the opportunity to design and built new GUI for presenting the tasks to participants. No research project is done without issues likewise the researcher also faced many issues during each phase of research life cycle. Thankfully with the great support and guidance from supervisor Dr Palaniappan Ramasamy, who got extensive knowledge in the field, the researcher was able to overcome the issues and finish the research in time. This project provided with very challenging tasks and taught the responsibilities in research work ethics and the researcher became passionate about it. During the period of this research, the researcher could experience formal and informal learning through enquiry research, reading and reflection. The confidence of researcher has been improved tremendously and she can transfer this knowledge to any other application/scenario in future.

CHAPTER 8

Conclusion

This research is a comprehensive study and one of the first attempt to completely analyse different bands and different features for different TDL for program comprehension tasks. From the results obtained, it can be concluded that there is definitely differences in TDL in EEG signal, but underlined pattern based on features are random. The classification result shows higher classification accuracy rate for easy and difficult tasks. So definitely there should some pattern that could be observed with advanced techniques, but too complex for the scope of this work.

Thus the researcher concludes that with a positive view that this research would help in better understanding of task difficulty level for program comprehension tasks.

8.1 Suggestion for further research

Even though noise like eye blink was removed from the EEG raw data, this research was limited by noise because of subject's physical movement and vocalization. This noise prevents the accuracy of the results, which are known to reflect the cognitive process of task difficulty level of program comprehension tasks. In further research, it is suggested to remove this unwanted and unknown noise as suggested in the study done by Repovs [47], which in-turn improves the accuracy of the results.

Instead of using all the features for classification, feature selection process could be implemented to eliminate the non-predictive features and avoid over-fitting of data. The study done by Lee and Tan [33] applied Weka's *CfsSubsetEval* operator to select features which are highly correlated with output variables and lower inter-correlation between features for classification. The same method could be adapted in further research to improve the accuracy of classification. More studies related to feature selection for non-linear EEG signals can be adapted in further research [19] [48]. In another study done by Palaniappan et al. [41] used genetic algorithm to select minimum number of channels to maximise the classification performance and achieved accuracy of 94.30% using ANN. The same algorithm could be adapted to improve the classifier accuracy.

References

[1] Turky Alotaiby, Fathi E Abd El-Samie, Saleh A Alshebeili, and Ishtiaq Ahmad. A review of channel selection algorithms for EEG signal processing. *EURASIP Journal on Advances in Signal Processing*, 2015(1):66, 2015.

[2] E W Anderson, K C Potter, L E Matzen, J F Shepherd, G a. Preston, and C T Silva. A user study of visualization effectiveness using EEG and cognitive load. *Eurographics '11*, 30(3):791–800, 2011.

[3] Pavlo D. Antonenko and Dale S. Niederhauser. The influence of leads on cognitive load and learning in a hypertext environment. *Computers in Human Behavior*, 26(2):140–150, 2010.

[4] Chris Berka, Daniel J Levendowski, Michelle N Lumicao, Alan Yau, Gene Davis, Vladimir T Zivkovic, Richard E Olmstead, Patrice D Tremoulet, and Patrick L Craven. Eeg correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine*, 78(Supplement 1):B231–B244, 2007.

[5] Jeffrey B Brookings, Glenn F Wilson, and Carolyne R Swain. Psychophysiological responses to changes in workload during simulated air traffic control. *Biological*

psychology, 42(3):361–377, 1996.

[6] Ruven Brooks. Towards a theory of the comprehension of computer programs. *International Journal of Man-Machine Studies*, 18(6):543–554, 1983.

[7] Roland Brünken, Jan L Plass, and Detlev Leutner. Direct Measurement of Cognitive Load in Multimedia Learning. *Educational Psychologist*, 38(1):53–61, 2003.

[8] Roland Brünken, Jan L Plass, and Detlev Leutner. Assessment of cognitive load in multimedia learning with dual-task methodology: Auditory load and modality effects. *Instructional Science*, 32(1-2):115–132, 2004.

[9] R Cabeza and L Nyberg. Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of cognitive neuroscience*, 12(1):1–47, 2000.

[10] Gabriele Cierniak, Katharina Scheiter, and Peter Gerjets. Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior*, 25(2):315–324, 2009.

[11] Igor Crk, Timothy Kluthe, and Andreas Stefik. Understanding Programming Expertise: An Empirical Study of Phasic Brain Wave Changes. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 23(1):2, 2015.

[12] Lionel E Deimel and J Fernando Naveda. Reading Computer Programs : Instructor ' s Guide and Exercises. *Order A Journal On The Theory Of Ordered Sets And Its Applications*, (August), 1990.

[13] Krista E. DeLeeuw and Richard E. Mayer. A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology*, 100(1):223–234, 2008.

[14] Emotive.com. EMOTIV Epoc - 14 Channel Wireless EEG Headset (2016) emotiv epoc - 14 channel wireless eeg headset (2016), 2016.

[15] K. Anders Ericsson and Herbert A. Simon. Verbal Reports as Data. *Psychological Review*, 87(3):215–251, 1980.

[16] eterlogic.com. Virtual Serial Port Emulator, year = 2016, url = http://www.eterlogic.com/Products.VSPE.html, urldate = 2016-09-02.

[17] Elizabeth a. Felton, Justin C. Williams, Gregg C. Vanderheiden, and Robert G. Radwin. Mental workload during brainâĂŞcomputer interface training. *Ergonomics*, 55(5):526–537, 2012.

[18] Thomas Fritz, Andrew Begel, Sebastian C Müller, Serap Yigit-Elliott, and Manuela Züger. Using Psycho-physiological Measures to Assess Task Difficulty in Software Development. *Proceedings of the 36th International Conference on Software Engineering*, pages 402–413, 2014.

[19] Deon Garrett, David A. Peterson, Charles W. Anderson, and Michael H. Thaut. Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2):141–144, 2003.

[20] Peter Gerjets, Carina Walter, Wolfgang Rosenstiel, Martin Bogdan, and Thorsten O. Zander. Cognitive state monitoring and the design of adaptive instruction in digital environments: Lessons learned from cognitive workload assessment using a passive brain-computer interface approach. *Frontiers in Neuroscience*, 8(DEC):1–21, 2014.

[21] D J Gilmore and T R G Green. Programming plans and programming expertise. *The Quarterly Journal of Experimental Psychology Section A*, 40(3):423–442, 1988.

[22] David Grimes, Desney S. Tan, Scott E. Hudson, Pradeep Shenoy, and Rajesh P.N. Rao. Feasibility and pragmatics of classifying working memory load with an electroencephalograph. *Proceedings of the 26th SIGCHI Conference on Human Factors in Computing Systems*, page 835, 2008.

[23] M. E. Hansen, a. Lumsdaine, and R. L. Goldstone. An experiment on the cognitive complexity of code. *Proceedings of the Thirty-Fifth Annual Conference of the Cognitive Science Society*, 2013.

[24] Sandra G. Hart and Lowell E. Staveland. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, 52(C):139–183, 1988.

[25] E Haselsteiner and G Pfurtscheller. Using time-dependent neural networks for EEG classification. *IEEE transactions on rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, 8(4):457–463, 2000.

[26] Lucy A. Henry. The Working Memory Model. *The Development of Working Memory in Children*, pages 1–36, 2011.

[27] Nai-jen Huan and Ramaswamy Palaniappan. BRAIN COMPUTER INTERFACE DESIGN USING MENTAL TASK CLASSIFICATION. pages 1–9, 2000.

[28] H H Jasper. The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology*, 10(2):371–375, 1958.

[29] Ole Jensen and Claudia D Tesche. Frontal theta activity in humans increases with memory load in a working memory task. *European journal of Neuroscience*, 15(8):1395–1399, 2002.

[30] W. Klimesch, M. Doppelmayr, Th Pachinger, and B. Ripper. Brain oscillations and human memory: EEG correlates in the upper alpha and theta band. *Neuroscience Letters*, 238(1-2):9–12, 1997.

[31] Wolfgang Klimesch. EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews*, 29(2-3):169–195, 1999.

[32] Wolfgang Klimesch. Alpha-band oscillations, attention, and controlled access to stored information. *Trends in cognitive sciences*, 16(12):606–617, 2012.

[33] J C Lee and D S Tan. Using a low-cost electroencephalograph for task classification in HCI research. *Proceedings of the 19th ACM Symposium on User Interface Software and Technology*, pages 81–90, 2006.

[34] Abhimanyu Malik, Samaksh Kumar, and Vijay Khare. Component Analysis And Neural Network Mental Task Classification Based on Principal Component Analysis and Neural Network. (August), 2016.

[35] Robert Malmivuo, Jaakko and Plonsey. *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields.* Oxford University Press, USA, 1995.

[36] Plamen Manoilov. EEG eye-blinking artefacts power spectrum analysis. *... of International Conference on Computer Systems ...*, pages 1–5, 2006.

[37] Loredana Mihalca, R. J C M Salden, Gemma Corbalan, Fred Paas, and Mircea Miclea. Effectiveness of cognitive-load based adaptive instruction in genetics education. *Computers in Human Behavior*, 27(1):82–88, 2011.

[38] H. Müller, K. Wong, and S. Tilley. Understanding software systems using reverse engineering technology. *The 62nd Congress of L'Association Canadienne Francaise pour l'Avancement des Sciences Proceedings (ACFAS)*, 26(4):41–48, 1994.

[39] Yi L. Murphey, Hong Guo, and A. Feldkamp Lee. Neural learning from unbalanced data. *Applied Intelligence*, 21(2 SPEC. ISS.):117–128, 2004.

[40] Takao Nakagawa, Yasutaka Kamei, Hidetake Uwano, Akito Monden, Kenichi Matsumoto, and Daniel M. German. Quantifying programmers' mental workload during program comprehension based on cerebral blood flow measurement: a controlled experiment. *Companion Proceedings of the 36th International Conference on Software Engineering - ICSE Companion 2014*, pages 448–451, 2014.

[41] R. Palaniappan, P. Raveendran, and S. Omatu. VEP optimal channel selection using genetic algorithm for neural network classification of alcoholics. *IEEE Transactions on Neural Networks*, 13(2):486–491, 2002.

[42] Ramaswamy Palaniappan. *Biological Signal Analysis.* 2010.

[43] Nancy Pennington. Stimulus structures and mental representations in expert comprehension of computer programs. *Cognitive Psychology*, 19(3):295–341, 1987.

[44] G Pfurtscheller and A Aranibar. Event-related cortical desynchronization detected by power measurements of scalp EEG. *Electroencephalography and clinical*, 1977.

[45] Alan T. Pope, Frederick G. Freeman, Mark W. Scerbo, and Peter J. Mikulka. Empirical Analysis of EEG and ERPs for Psychophysiological Adaptive Task Allocation. (June), 2001.

[46] David Reby, Sovan Lek, Ioannis Dimopoulos, Jean Joachim, Jacques Lauga, and Stéphane Aulagnier. Artificial neural networks as a classification method in the behavioural sciences. *Behavioural Processes*, 40(1):35–43, 1997.

[47] Grega Repovš. Dealing with Noise in EEG Recording and Data Analysis. *Informatica Medica Slovenica*, 15(1):18–25, 2010.

[48] Germán Rodríguez-Bermúdez, Pedro J. García-Laencina, Joaquín Roca-González, and Joaquín Roca-Dorda. Efficient feature selection and linear discrimination of EEG signals. *Neurocomputing*, 115:161–165, 2013.

[49] Dale Shaffer, Wendy Doube, and Juhani E. Tuovinen. Applying Cognitive Load Theory to Computer Science Education. *Ppig '03*, (April):333–346, 2003.

[50] Bonita Sharif and Jonathan I. Maletic. An eye tracking study on camel case and under score identifier styles. *Program Comprehension (ICPC), 2012 IEEE 20th International Conference on*, 2012.

[51] Janet Siegmund, Christian KÂÍAstner, JÂÍorg Liebig, Sven Apel, and Stefan Hanenberg. Measuring and modeling programming experience. *Empirical Software Engineering*, 19(5):1299–1334, 2014.

[52] Raymond Carl Smith. Smith, Raymond Carl. "Electroencephalograph based brain computer interfaces." PhD diss., University College Dublin, 2004. *Engineering*, 2004.

[53] Kate Soloway, Elliot and Ehrlich. Empirical studies of programming knowledge. *IEEE Transactions on Software Engineering*, pages 595—-609, 1984.

[54] Margaret Anne Storey. Theories, tools and research methods in program comprehension: Past, present and future. *Software Quality Journal*, 14(3):187–208, 2006.

[55] Michal Teplan. Fundamentals of EEG measurement. *Measurement Science Review*, 2:1–11, 2002.

[56] Ting Wang, Sheng-uei Guan, Ka Lok Man, and T O Ting. EEG Eye State Identification Using Incremental Attribute Learning with Time-Series Classification. 2014(April), 2014.

[57] Glenn F. Wilson and Christopher a. Russell. Real-Time Assessment of Mental Workload Using Psychophysiological Measures and Artificial Neural Networks. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 45(4):635–643, 2003.

[58] Muhammed Yousoof and Mohd Sapiyan. Cognitive Load Measurement in Learning Programming Using NASA TLX rating scale (Non Physiological Measures). pages 235–245.

[59] Pega Zarjam, Julien Epps, and Fang Chen. Characterizing working memory load using eeg delta activity. (Eusipco 2011):1554–1558, 2015.

APPENDIX A

EEG Data - Subject based feature results (plot)

Below gives the diagrammatic representation of results of computed Energy, computed ERD and computed Frequency ratio for each subject based on TDL. The X-axis coordinates are represented as (1-Energy lower 1 alpha, 2-Energy lower 2 Alpha, 3-Energy upper Alpha, 4-Energy Theta(IAF), 5-Energy Delta, 6-Energy Alpha, 7-Energy Beta, 8-Energy Theta, 9-Energy Gamma, 10-Energy (alpha + theta), 11-Energy (alpha + beta), 12-Energy all frequencies, 13-%ERD lower 1 alpha, 14-%ERD lower 2 Alpha, 15-%ERD upper Alpha, 16-%ERD Theta(IAF), 17-%ERD Delta, 18- %ERD Alpha, 19-%ERD Beta, 20-%ERD Theta, 21-%ERD Gamma, 22-%ERD (alpha + theta), 23-%ERD (alpha + beta), 24-%ERD all frequencies, 25-($\Delta(\alpha/\beta)$), 26-($\Delta(\alpha/\gamma)$),27-($\Delta(\alpha/\delta)$),28-($\Delta(\alpha/\theta)$),29-($\Delta(\beta/\alpha)$),30-($\Delta(\beta/\gamma)$), 31-($\Delta(\beta/\delta)$), 32 -($\Delta(\beta/\theta)$), 33-($\Delta(\gamma/\alpha)$),34-($\Delta(\gamma/\beta)$), 35-($\Delta(\gamma/\delta)$), 36-($\Delta(\gamma/\theta)$),37-($\Delta(\delta/\alpha)$),38-($\Delta(\delta/\beta)$),39-($\Delta(\delta/\gamma)$),40-($\Delta(\delta/\theta)$),41-($\Delta(\theta/\alpha)$), 42-($\Delta(\theta/\beta)$), 43-($\Delta(\theta/\gamma)$), 44-($\Delta(\theta/\delta)$),45-($\Delta(\theta/(\alpha+\beta))$),46-($\Delta(\beta/(\alpha+\theta))$))

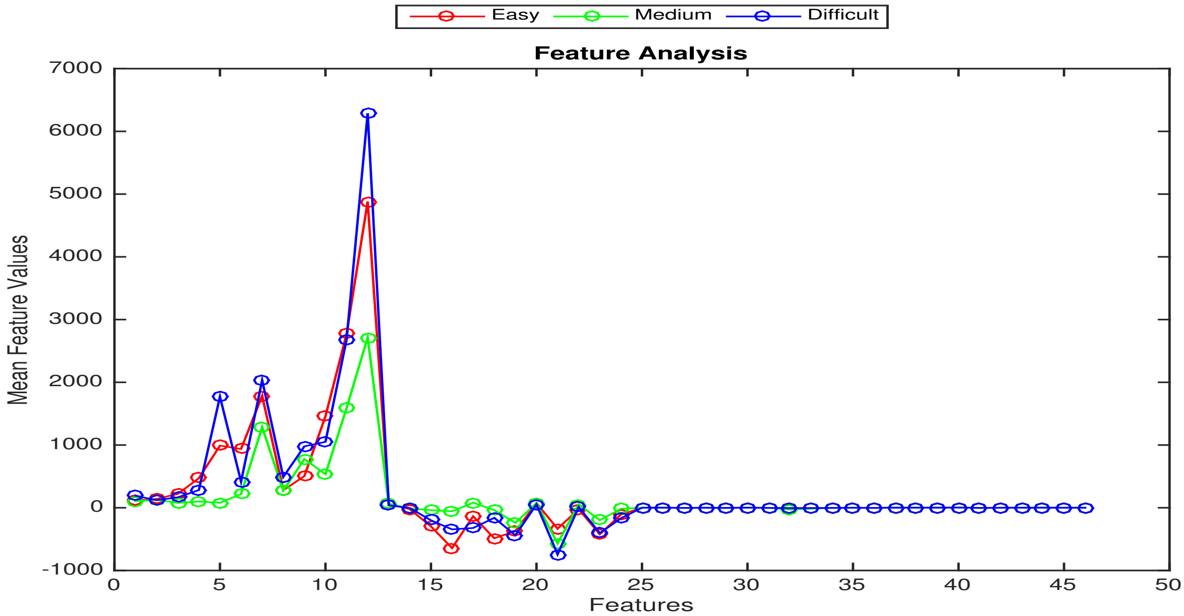


Figure A.1: Feature analysis - Subject ID - 1

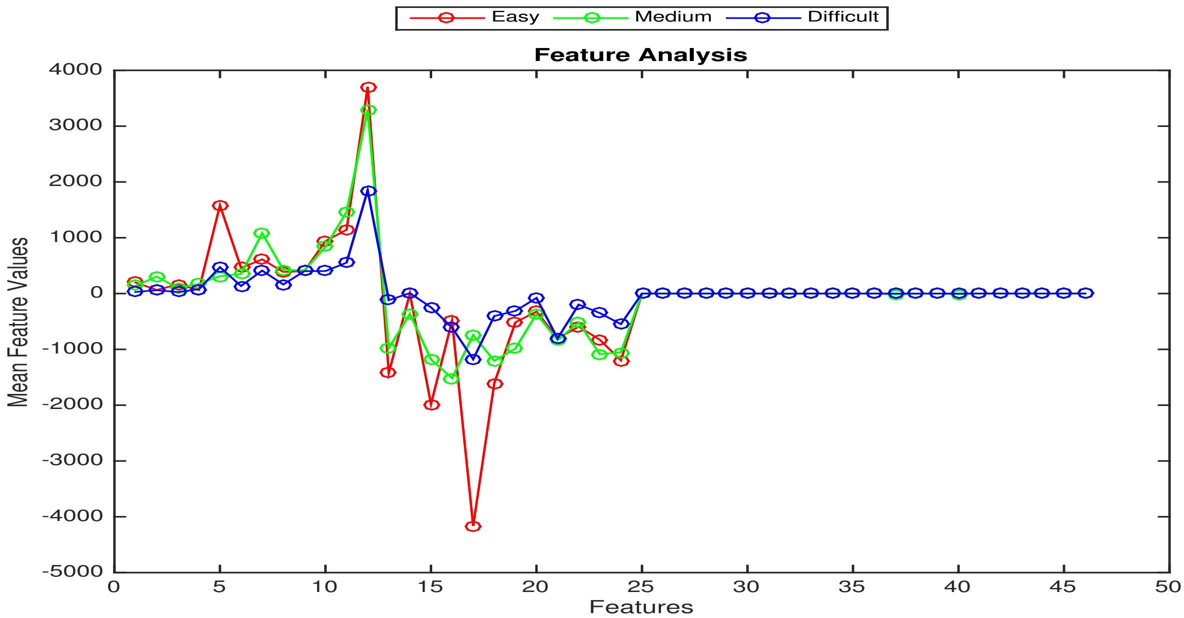


Figure A.2: Feature analysis - Subject ID - 3

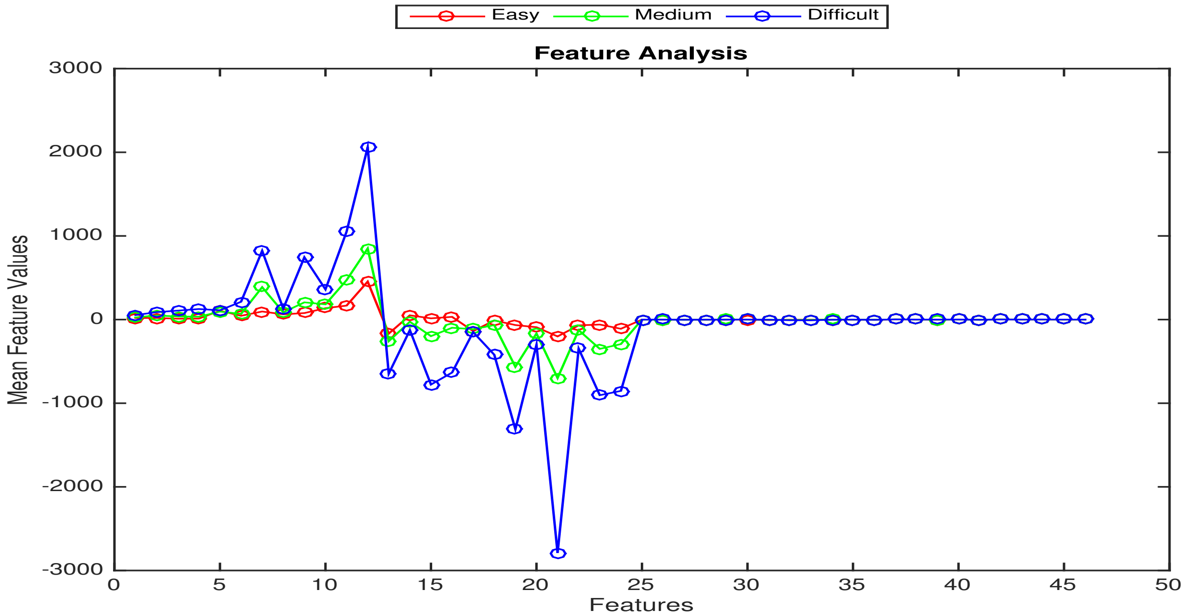


Figure A.3: Feature analysis - Subject ID - 4

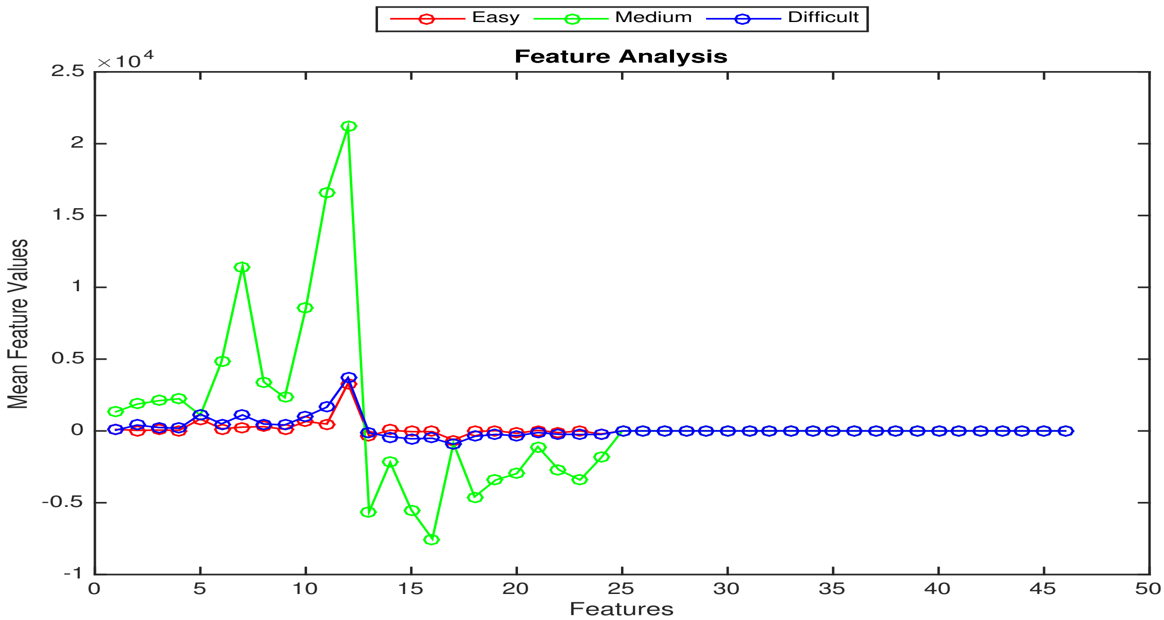


Figure A.4: Feature analysis - Subject ID - 5

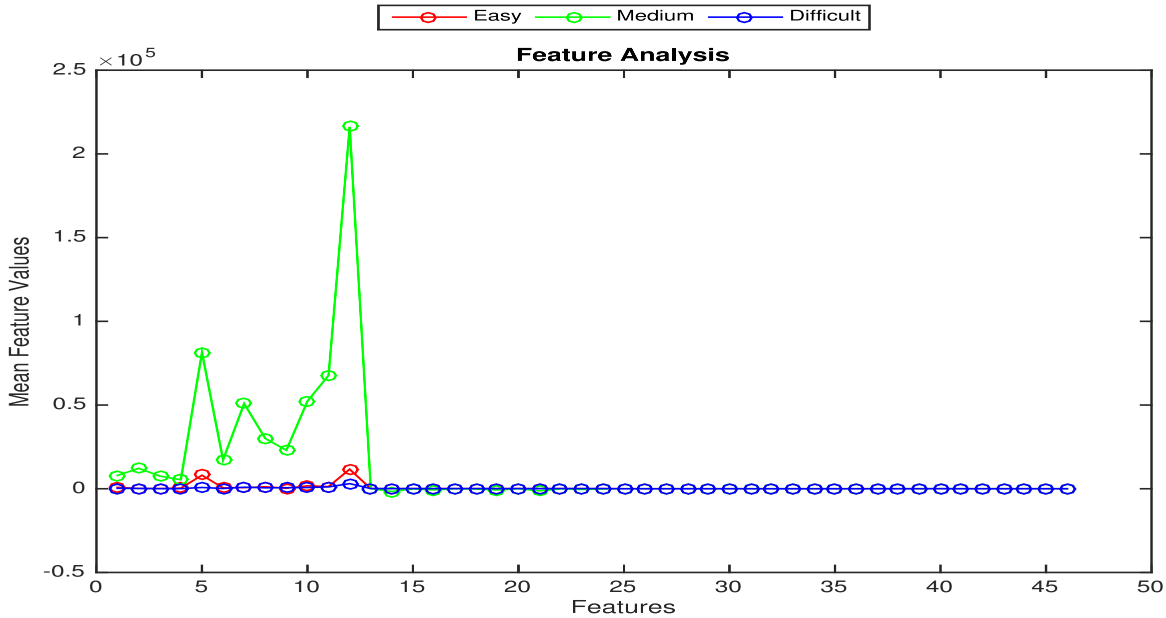


Figure A.5: Feature analysis - Subject ID - 6

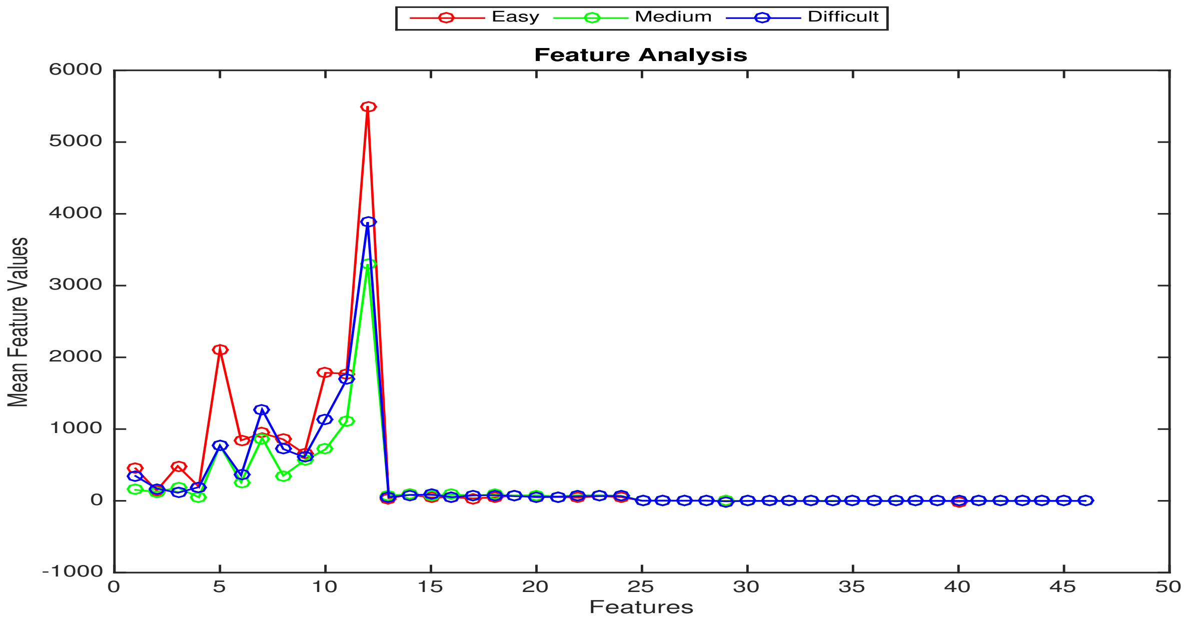


Figure A.6: Feature analysis - Subject ID - 8

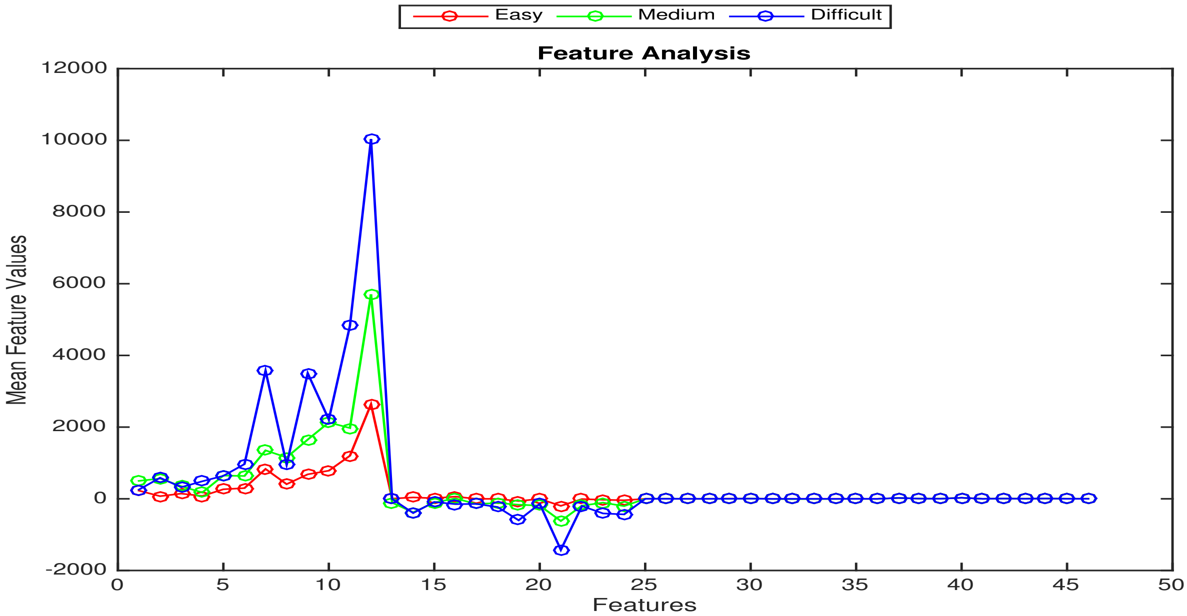


Figure A.7: Feature analysis - Subject ID - 9

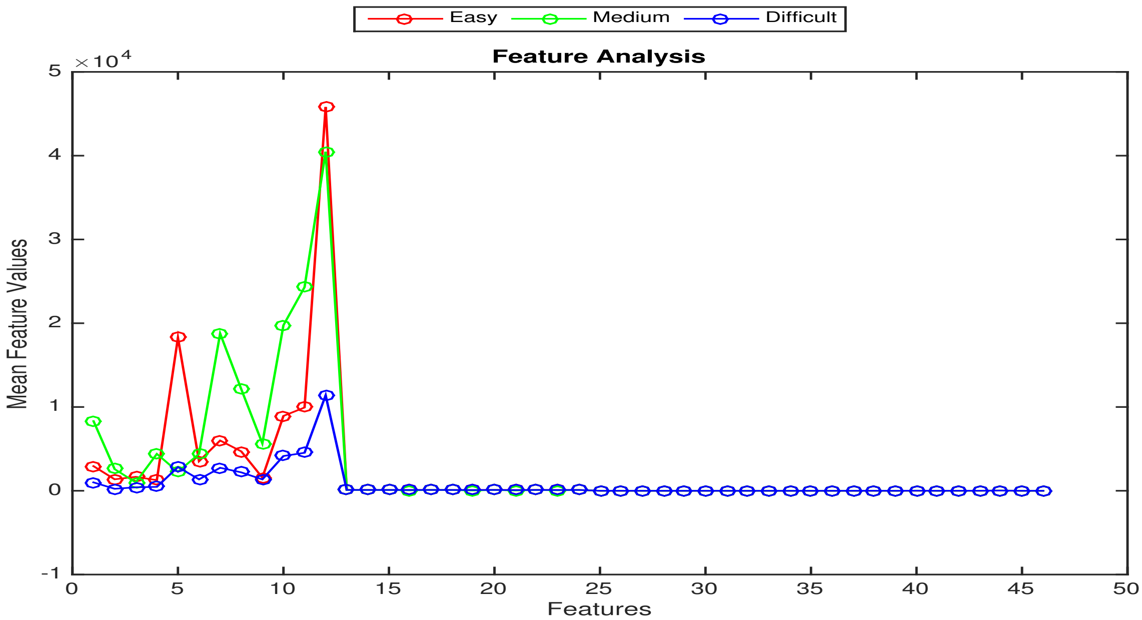


Figure A.8: Feature analysis - Subject ID - 10

APPENDIX B

Subject based energy feature and statistical analysis results

In this chapter, the results of computer mean Energy and Kurskal-Wallis test are presented for each subject.

Table B.1: Energy Analysis - SubjectID 1

Feature name	**Mean Energy**			**P Value**	**df**	**Chi-Squ**	**No. of samples**
	Easy	Medium	Difficult				
Lower 1 Alpha	112.09	105.81	190.88	4.50E-11	2	47.64	(255, 297, 338)
Lower 2 Alpha	142.74	126.78	120.38	7.55E-08	2	32.79	(255, 297, 338)
Upper Alpha	220.23	77.15	166.49	1.06E-26	2	119.62	(255, 297, 338)
Theta (IAF)	471.21	98.81	281.82	3.55E-34	2	154.04	(255, 297, 338)
Delta	989.96	79.22	1770.76	4.08E-100	2	457.70	(255, 297, 338)
Alpha	937.86	218.88	416.4	5.07E-49	2	222.40	(255, 297, 338)
Beta	1786.15	1286.11	2033.43	3.64E-20	2	89.52	(255, 297, 338)
Theta	279.97	273.51	491.07	2.85E-08	2	34.74	(255, 297, 338)
Gamma	504.17	776.15	979.14	6.29E-51	2	231.18	(255, 297, 338)
Alpha + Theta	1459.41	535.39	1050.82	7.44E-57	2	258.4	(255, 297, 338)
Alpha + Beta	2768.91	1588.08	2678.79	1.37E-24	2	109.89	(255, 297, 338)
All frequencies	4880.14	2715.76	6287.65	3.02E-41	2	186.60	(255, 297, 338)

Table B.2: Energy Analysis - SubjectID 3

Feature name	**Mean Energy**			**P Value**	**df**	**Chi-Squ**	**No. of samples**
	Easy	Medium	Difficult				
Lower 1 Alpha	198.26	140.25	28.56	2.15E-44	2	201.09	(105, 205, 263)
Lower 2 Alpha	65.07	305.02	59.67	1.01E-21	2	96.68	(105, 205, 263)
Upper Alpha	161.29	99.15	26.81	1.92E-16	2	72.37	(105, 205, 263)
Theta (IAF)	63.41	174.81	76.76	1.49E-16	2	72.8	(105, 205, 263)
Delta	1580.43	312.22	473.08	2.98E-17	2	76.10	(105, 205, 263)
Alpha	458.15	350.39	135.01	2.45E-15	2	67.28	(105, 205, 263)
Beta	614.28	1083.16	411.85	3.91E-28	2	126.21	(105, 205, 263)
Theta	387.08	425.25	162.85	3.73E-22	2	98.67	(105, 205, 263)
Gamma	400.84	415.33	401.59	9.78E-05	2	18.46	(105, 205, 263)
Alpha + Theta	941.23	836.68	403.25	6.10E-14	2	60.85	(105, 205, 263)
Alpha + Beta	1142.34	1462.85	553.51	5.61E-24	2	107.07	(105, 205, 263)
All frequencies	3704.74	3293.33	1851.47	4.78E-19	2	84.37	(105, 205, 263)

Table B.3: Energy Analysis - SubjectID 4

Feature name	**Energy Mean**			**P Value**	**df**	**Chi-Squ**	**No. of samples**
	Easy	Medium	Difficult				
Lower 1 Alpha	19.32	25.85	53.91	7.26E-07	2	28.27	(277, 365, 539)
Lower 2 Alpha	19.29	49.59	88.76	5.07E-32	2	144.12	(277, 365, 539)
Upper Alpha	10.44	36.43	107.11	1.05E-25	2	115.03	(277, 365, 539)
Theta (IAF)	11.62	33.96	125.99	6.90E-31	2	138.90	(277, 365, 539)
Delta	105.95	95.61	111.13	0.053639	2	5.85	(277, 365, 539)
Alpha	45.01	67.92	212.74	4.51E-07	2	29.22	(277, 365, 539)
Beta	96.46	387.37	822.51	2.96E-70	2	320.19	(277, 365, 539)
Theta	62.30	86.75	130.61	3.47E-28	2	126.45	(277, 365, 539)
Gamma	77.72	205.86	744.95	7.85E-56	2	253.77	(277, 365, 539)
Alpha + Theta	136.15	182.13	360.85	1.38E-15	2	68.43	(277, 365, 539)
Alpha + Beta	166.90	471.12	1050.75	5.04E-35	2	157.95	(277, 365, 539)
All frequencies	454.08	849.48	2058.64	1.16E-19	2	87.19	(277, 365, 539)

Table B.4: Energy Analysis - SubjectID 5

Feature name	**Mean Energy**			**P Value**	**df**	**Chi-Squ**	**No. of samples**
	Easy	Medium	Difficult				
Lower 1 Alpha	103.67	1335.41	56.05	2.34E-16	2	71.98	(113, 224, 291)
Lower 2 Alpha	39.97	1867.34	409.73	7.78E-11	2	46.55	(113, 224, 291)
Upper Alpha	57.03	2100.28	243.26	7.32E-07	2	28.25	(113, 224, 291)
Theta (IAF)	43.17	2241.68	179.88	1.52E-07	2	31.40	(113, 224, 291)
Delta	829.56	1089.98	1164.05	1.65E-06	2	26.63	(113, 224, 291)
Alpha	151.71	4813.94	490.38	7.03E-06	2	23.73	(113, 224, 291)
Beta	266.94	11451.16	1112.15	4.01E-22	2	98.54	(113, 224, 291)
Theta	297.35	3418.41	482.39	0.00071733	2	14.48	(113, 224, 291)
Gamma	113.70	2324.32	416.53	7.61E-40	2	180.15	(113, 224, 291)
Alpha + Theta	654.55	8536.15	1008.19	0.045442	2	6.18	(113, 224, 291)
Alpha + Beta	481.45	16628.65	1658.17	9.33E-12	2	50.80	(113, 224, 291)
All frequencies	3310.22	21171.08	3702.23	0.0010219	2	13.77	(113, 224, 291)

Table B.5: Energy Analysis - SubjectID 6

Feature name	**Mean Energy**			**P Value**	**df**	**Chi-Squ**	**No. of samples**
	Easy	Medium	Difficult				
	837.70	7940.92	250.76	1.75E-22	2	100.19	(106, 236, 629)
Lower 2 Alpha	221.97	12213.87	167.26	1.27E-15	2	68.60	(106, 236, 629)
Upper Alpha	125.86	7584.51	44.46	6.12E-27	2	120.72	(106, 236, 629)
Theta (IAF)	389.71	5365.96	65.91	2.06E-23	2	104.48	(106, 236, 629)
Delta	8067.78	81431.57	766.79	3.34E-27	2	121.93	(106, 236, 629)
Alpha	558.13	17257.32	143.36	6.28E-22	2	97.64	(106, 236, 629)
Beta	662.18	50859.84	912.50	3.29E-14	2	62.09	(106, 236, 629)
Theta	1204.22	30043.47	563.33	7.75E-12	2	51.17	(106, 236, 629)
Gamma	273.95	22989.61	610.94	1.01E-18	2	82.88	(106, 236, 629)
Alpha + Theta	1785.11	52331.30	736.44	1.45E-12	2	54.53	(106, 236, 629)
Alpha + Beta	1278.60	67721.34	1099.33	6.49E-14	2	60.73	(106, 236, 629)
All frequencies	11635.03	216164.85	3059.88	1.36E-08	2	36.23	(106, 236, 629)

Table B.6: Energy Analysis - SubjectID 8

Feature name	Mean Energy			P Value	df	Chi-Squ	No. of samp
	Easy	Medium	Difficult				
Lower 1 Alpha	452.55	152.94	346.17	4.43E-16	2	70.71	(93, 186, 225
Lower 2 Alpha	143.31	126.78	171.99	0.061461	2	5.58	(93, 186, 225
Upper Alpha	484.75	184.10	112.90	1.19E-07	2	31.90	(93, 186, 225
Theta (IAF)	196.43	50.29	191.31	0.00016766	2	17.39	(93, 186, 225
Delta	2110.21	773.02	768.48	4.13E-16	2	70.85	(93, 186, 225
Alpha	842.71	259.36	356.81	1.61E-07	2	31.29	(93, 186, 225
Beta	944.17	873.41	1270.56	0.29884	2	2.42	(93, 186, 225
Theta	856.48	351.82	721.96	2.36E-15	2	67.36	(93, 186, 225
Gamma	658.51	560.92	608.57	4.17E-06	2	24.77	(93, 186, 225
Alpha + Theta	1784.21	718.51	1142.37	9.71E-17	2	73.74	(93, 186, 225
Alpha + Beta	1773.60	1105.38	1689.96	0.0030547	2	11.58	(93, 186, 225
All frequencies	5499.58	3298.01	3883.57	1.75E-07	2	31.12	(93, 186, 225

Table B.7: Energy Analysis - SubjectID 9

Feature name	Mean Energy			P Value	df	Chi-Squ	No. of samples
	Easy	Medium	Difficult				
Lower 1 Alpha	227.46	500.19	256.33	1.32E-28	2	128.39	(280, 460, 521)
Lower 2 Alpha	72.96	564.79	586.22	1.69E-39	2	178.55	(280, 460, 521)
Upper Alpha	144.84	352.20	323.65	2.76E-31	2	140.73	(280, 460, 521)
Theta (IAF)	77.42	175.07	487.99	7.90E-25	2	111.00	(280, 460, 521)
Delta	265.89	648.57	640.41	2.72E-25	2	113.12	(280, 460, 521)
Alpha	297.69	635.99	966.89	1.31E-32	2	146.83	(280, 460, 521)
Beta	831.94	1351.16	3570.39	6.42E-09	2	37.73	(280, 460, 521)
Theta	408.24	1154.19	955.92	1.27E-31	2	142.28	(280, 460, 521)
Gamma	682.89	1652.18	3502.85	0.012362	2	8.79	(280, 460, 521)
Alpha + Theta	784.33	2131.28	2210.87	5.05E-37	2	167.15	(280, 460, 521)
Alpha + Beta	1201.20	1973.82	4822.86	1.79E-27	2	123.18	(280, 460, 521)
All frequencies	2637.86	5703.09	10033.02	1.25E-08	2	36.40	(280, 460, 521)

Table B.8: Energy Analysis - SubjectID 10

Feature name	**Mean Energy**			**P Value**	**df**	**Chi-Squ**	**No. of samples**
	Easy	Medium	Difficult				
Lower 1 Alpha	2938.13	8318.10	1001.66	7.60E-31	2	138.70	(264, 699, 429)
Lower 2 Alpha	1369.53	2584.68	236.36	2.12E-49	2	224.16	(264, 699, 429)
Upper Alpha	1663.99	995.66	415.99	7.13E-44	2	198.70	(264, 699, 429)
Theta (IAF)	1267.79	4401.25	464.90	2.83E-16	2	71.60	(264, 699, 429)
Delta	18262.33	2261.34	2833.40	7.17E-35	2	157.24	(264, 699, 429)
Alpha	3531.92	4404.60	1360.69	8.57E-25	2	110.83	(264, 699, 429)
Beta	5996.69	18771.62	2745.09	1.03E-27	2	124.27	(264, 699, 429)
Theta	4704.59	12136.79	2197.76	2.22E-05	2	21.43	(264, 699, 429)
Gamma	1571.20	5651.75	1358.30	6.94E-91	2	415.20	(264, 699, 429)
Alpha + Theta	8935.43	19759.92	4149.67	0.0013669	2	13.19	(264, 699, 429)
Alpha + Beta	9939.25	24303.68	4533.87	1.67E-36	2	164.76	(264, 699, 429)
All frequencies	45811.84	40479.33	11396.32	1.67E-24	2	109.50	(264, 699, 429)

APPENDIX C

Subject based ERD feature and statistical analysis results

In this chapter, the results of computer percentage mean ERD and Kurskal-Wallis test are presented for each subject.

Table C.1: Percentage ERD Analysis - SubjectID 1

Feature name	**% Mean ERD**			**P Value**	**df**	**Chi-Squ**	**No. of samples**
	Easy	Medium	Difficult				
Lower 1 Alpha	63.79	65.82	38.33	4.50E-11	2	47.65	(255, 297, 338)
Lower 2 Alpha	-23.83	-9.99	-4.43	7.55E-08	2	32.80	(255, 297, 338)
Upper Alpha	-296.70	-38.97	-199.90	1.06E-26	2	119.62	(255, 297, 338)
Theta (IAF)	-643.17	-55.84	-344.47	3.55E-34	2	154.04	(255, 297, 338)
Delta	-136.18	81.10	-322.45	4.08E-100	2	457.70	(255, 297, 338)
Alpha	-486.51	-36.88	-160.46	5.07E-49	2	222.41	(255, 297, 338)
Beta	-373.21	-240.74	-438.73	3.64E-20	2	89.52	(255, 297, 338)
Theta	73.55	74.17	53.61	2.85E-08	2	34.75	(255, 297, 338)
Gamma	-334.12	-568.32	-743.11	6.29E-51	2	231.19	(255, 297, 338)
Alpha + Theta	-22.55	55.04	11.76	7.44E-57	2	258.48	(255, 297, 338)
Alpha + Beta	-421.97	-199.37	-404.98	1.37E-24	2	109.89	(255, 297, 338)
All frequencies	-108.35	-15.94	-168.44	3.02E-41	2	186.60	(255, 297, 338)

Table C.2: Percentage ERD Analysis - SubjectID 3

Feature name	**%Mean ERD**			**P Value**	**df**	**Chi-Squ**	**No. of samples**
	Easy	Medium	Difficult				
Lower 1 Alpha	-1424.22	-978.26	-119.58	2.15E-44	2	201.09	(105, 205, 263)
Lower 2 Alpha	0.72	-365.36	8.96	1.01E-21	2	96.69	(105, 205, 263)
Upper Alpha	-1981.64	-1179.70	-246.01	1.92E-16	2	72.37	(105, 205, 263)
Theta (IAF)	-486.40	-1516.48	-609.85	1.49E-16	2	72.89	(105, 205, 263)
Delta	-4164.46	-742.48	-1176.52	2.98E-17	2	76.10	(105, 205, 263)
Alpha	-1606.65	-1205.26	-402.95	2.45E-15	2	67.29	(105, 205, 263)
Beta	-515.56	-985.41	-312.71	3.91E-28	2	126.22	(105, 205, 263)
Theta	-322.89	-364.59	-77.92	3.73E-22	2	98.68	(105, 205, 263)
Gamma	-802.55	-835.18	-804.24	9.78E-05	2	18.46	(105, 205, 263)
Alpha + Theta	-594.14	-517.03	-197.39	6.10E-14	2	60.86	(105, 205, 263)
Alpha + Beta	-827.00	-1087.09	-349.17	5.61E-24	2	107.08	(105, 205, 263)
All frequencies	-1197.22	-1053.17	-548.30	4.78E-19	2	84.37	(105, 205, 263)

Table C.3: Percentage ERD Analysis - SubjectID 4

Feature name	**% Mean ERD**			**P Value**	**df**	**Chi-Squ**	**No. of samples**
	Easy	Medium	Difficult				
Lower 1 Alpha	-167.58	-257.98	-646.59	7.26E-07	2	28.27	(277, 365, 539)
Lower 2 Alpha	50.85	-26.33	-126.14	5.07E-32	2	144.12	(277, 365, 539)
Upper Alpha	14.53	-198.22	-776.87	1.05E-25	2	115.03	(277, 365, 539)
Theta (IAF)	32.35	-97.72	-633.55	6.90E-31	2	138.90	(277, 365, 539)
Delta	-138.29	-115.05	-149.94	0.053639	2	5.85	(277, 365, 539)
Alpha	-10.10	-66.17	-420.44	4.51E-07	2	29.22	(277, 365, 539)
Beta	-65.54	-564.80	-1311.57	2.96E-70	2	320.19	(277, 365, 539)
Theta	-91.52	-166.66	-301.49	3.47E-28	2	126.45	(277, 365, 539)
Gamma	-201.49	-698.54	-2789.72	7.85E-56	2	253.77	(277, 365, 539)
Alpha + Theta	-69.15	-126.29	-348.34	1.38E-15	2	68.43	(277, 365, 539)
Alpha + Beta	-59.51	-350.27	-904.26	5.04E-35	2	157.95	(277, 365, 539)
All frequencies	-110.17	-293.17	-852.82	1.16E-19	2	87.19	(277, 365, 539)

Table C.4: Percentage ERD Analysis - SubjectID 5

Feature name	**% Mean ERD**			**P Value**	**df**	**Chi-Squ**	**No. of samples**
	Easy	Medium	Difficult				
Lower 1 Alpha	-344.93	-5631.18	-140.55	2.34E-16	2	71.98	(113, 224, 291)
Lower 2 Alpha	50.87	-2195.57	-403.69	7.78E-11	2	46.55	(113, 224, 291)
Upper Alpha	-52.01	-5497.78	-548.34	7.32E-07	2	28.25	(113, 224, 291)
Theta (IAF)	-46.68	-7517.53	-511.26	1.52E-07	2	31.40	(113, 224, 291)
Delta	-633.64	-863.94	-929.44	1.65E-06	2	26.63	(113, 224, 291)
Alpha	-47.70	-4586.42	-377.39	7.03E-06	2	23.73	(113, 224, 291)
Beta	17.98	-3418.53	-241.72	4.01E-22	2	98.54	(113, 224, 291)
Theta	-166.81	-2967.31	-332.84	0.00071733	2	14.48	(113, 224, 291)
Gamma	41.56	-1094.63	-114.08	7.61E-40	2	180.15	(113, 224, 291)
Alpha + Theta	-115.79	-2714.10	-232.37	0.045442	2	6.18	(113, 224, 291)
Alpha + Beta	-1.76	-3414.65	-250.47	9.33E-12	2	50.80	(113, 224, 291)
All frequencies	-207.10	-1864.08	-243.46	0.0010219	2	13.77	(113, 224, 291)

Table C.5: Percentage ERD Analysis - SubjectID 6

Feature name	**% Mean ERD**			**P Value**	**df**	**Chi-Squ**	**No. of samples**
	Easy	Medium	Difficult				
Lower 1 Alpha	95.71	59.34	98.72	1.75E-22	2	100.19	(106, 236, 629)
Lower 2 Alpha	68.28	-1645.15	76.10	1.27E-15	2	68.60	(106, 236, 629)
Upper Alpha	96.26	-125.56	98.68	6.12E-27	2	120.72	(106, 236, 629)
Theta (IAF)	44.55	-663.51	90.62	2.06E-23	2	104.48	(106, 236, 629)
Delta	83.03	-71.26	98.39	3.34E-27	2	121.93	(106, 236, 629)
Alpha	89.93	-211.30	97.41	6.28E-22	2	97.64	(106, 236, 629)
Beta	89.01	-744.39	84.85	3.29E-14	2	62.09	(106, 236, 629)
Theta	94.68	-32.65	97.51	7.75E-12	2	51.17	(106, 236, 629)
Gamma	83.44	-1290.07	63.06	1.01E-18	2	82.88	(106, 236, 629)
Alpha + Theta	93.52	-90.06	97.33	1.45E-12	2	54.53	(106, 236, 629)
Alpha + Beta	90.18	-420.28	91.55	6.49E-14	2	60.73	(106, 236, 629)
All frequencies	84.98	-179.03	96.05	1.36E-08	2	36.23	(106, 236, 629)

Table C.6: Percentage ERD Analysis - SubjectID 8

Feature name	**%Mean ERD**			**P Value**	**df**	**Chi-Squ**	**No. of samples**
	Easy	Medium	Difficult				
Lower 1 Alpha	36.88	78.67	51.72	4.43E-16	2	70.71	(93, 186, 225)
Lower 2 Alpha	84.79	86.54	81.74	0.061461	2	5.58	(93, 186, 225)
Upper Alpha	46.28	79.60	87.49	1.19E-07	2	31.90	(93, 186, 225)
Theta (IAF)	50.22	87.26	51.52	0.00016766	2	17.39	(93, 186, 225)
Delta	34.30	75.93	76.07	4.13E-16	2	70.85	(93, 186, 225)
Alpha	48.86	84.26	78.35	1.61E-07	2	31.29	(93, 186, 225)
Beta	73.18	75.19	63.91	0.29884	2	2.42	(93, 186, 225)
Theta	49.15	79.11	57.14	2.36E-15	2	67.36	(93, 186, 225)
Gamma	53.55	60.43	57.07	4.17E-06	2	24.77	(93, 186, 225)
Alpha + Theta	44.66	77.72	64.57	9.71E-17	2	73.74	(93, 186, 225)
Alpha + Beta	65.50	78.50	67.13	0.0030547	2	11.58	(93, 186, 225)
All frequencies	52.40	71.46	66.39	1.75E-07	2	31.12	(93, 186, 225)

Table C.7: Percentage ERD Analysis - SubjectID 9

Feature name	**% Mean ERD**			**P Value**	**df**	**Chi-Squ**	**No. of samples**
	Easy	Medium	Difficult				
Lower 1 Alpha	6.22	-106.24	-5.69	1.32E-28	2	128.39	(280, 460, 521)
Lower 2 Alpha	38.14	-378.91	-397.08	1.69E-39	2	178.55	(280, 460, 521)
Upper Alpha	15.71	-104.96	-88.35	2.76E-31	2	140.73	(280, 460, 521)
Theta (IAF)	60.58	10.86	-148.48	7.90E-25	2	111.00	(280, 460, 521)
Delta	-1.04	-146.46	-143.36	2.72E-25	2	113.12	(280, 460, 521)
Alpha	-1.31	-116.44	-229.05	1.31E-32	2	146.83	(280, 460, 521)
Beta	-62.92	-164.60	-599.19	6.42E-09	2	37.73	(280, 460, 521)
Theta	-2.36	-189.41	-139.69	1.27E-31	2	142.28	(280, 460, 521)
Gamma	-197.12	-618.84	-1424.05	0.012362	2	8.79	(280, 460, 521)
Alpha + Theta	-7.03	-190.82	-201.68	5.05E-37	2	167.15	(280, 460, 521)
Alpha + Beta	-26.55	-107.95	-408.09	1.79E-27	2	123.18	(280, 460, 521)
All frequencies	-40.67	-204.14	-435.04	1.25E-08	2	36.40	(280, 460, 521)

Table C.8: Percentage ERD Analysis - SubjectID 10

Feature name	**% Mean ERD**			**P Value**	**df**	**Chi-Squ**	**No. of samples**
	Easy	Medium	Difficult				
Lower 1 Alpha	96.99	91.48	98.97	7.60E-31	2	138.70	(264, 699, 429)
Lower 2 Alpha	92.78	86.37	98.75	2.12E-49	2	224.16	(264, 699, 429)
Upper Alpha	87.50	92.52	96.87	7.13E-44	2	198.70	(264, 699, 429)
Theta (IAF)	86.11	51.77	94.91	2.83E-16	2	71.60	(264, 699, 429)
Delta	89.78	98.73	98.41	7.17E-35	2	157.24	(264, 699, 429)
Alpha	87.31	84.18	95.11	8.57E-25	2	110.83	(264, 699, 429)
Beta	78.65	33.18	90.23	1.03E-27	2	124.27	(264, 699, 429)
Theta	97.97	94.76	99.05	2.22E-05	2	21.43	(264, 699, 429)
Gamma	80.85	31.10	83.44	6.94E-91	2	415.20	(264, 699, 429)
Alpha + Theta	96.57	92.42	98.41	0.0013669	2	13.19	(264, 699, 429)
Alpha + Beta	82.44	57.07	91.99	1.67E-36	2	164.76	(264, 699, 429)
All frequencies	91.04	92.08	97.77	1.67E-24	2	109.50	(264, 699, 429)

APPENDIX D

Subject based frequency ratio feature and statistical analysis results

In this chapter, the results of computer mean Frequency ratio and Kurskal-Wallis test are presented for each subject.

Table D.1: Frequency ratio Analysis - SubjectID 1

Feature name	**Mean Frequency Ratio**			**P Value**	**df**	**Chi-Squ**	**No. of samples**
	Easy	Medium	Difficult				
$\Delta(\alpha/\beta)$	0.09	0.27	0.18	3.41E-24	2	108.07	(255, 297, 338)
$\Delta(\alpha/\gamma)$	0.94	1.05	0.79	1.13E-12	2	55.01	(255, 297, 338)
$\Delta(\alpha/\delta)$	-2.06	-3.13	-1.91	1.33E-05	2	22.45	(255, 297, 338)
$\Delta(\alpha/\theta)$	-0.70	-0.16	-0.17	0.27665	2	2.57	(255, 297, 338)
$\Delta(\beta/\alpha)$	-1.23	-4.01	-1.85	3.41E-24	2	108.07	(255, 297, 338)
$\Delta(\beta/\gamma)$	1.46	1.11	0.91	0.002458	2	12.02	(255, 297, 338)
$\Delta(\beta/\delta)$	-1.77	-2.78	-1.75	3.68E-22	2	98.71	(255, 297, 338)
$\Delta(\beta/\theta)$	-4.51	-28.62	-6.66	5.65E-17	2	74.82	(255, 297, 338)
$\Delta(\gamma/\alpha)$	-3.80	-16.44	-5.69	9.97E-13	2	55.27	(255, 297, 338)
$\Delta(\gamma/\beta)$	-1.38	-3.11	-2.27	1.13E-12	2	55.01	(255, 297, 338)
$\Delta(\gamma/\delta)$	-0.37	-0.32	-0.37	0.002458	2	12.02	(255, 297, 338)
$\Delta(\gamma/\theta)$	-1.97	-2.38	-2.11	5.59E-08	2	33.40	(255, 297, 338)
$\Delta(\delta/\alpha)$	0.93	1.91	1.64	1.33E-05	2	22.45	(255, 297, 338)
$\Delta(\delta/\beta)$	0.90	1.14	1.01	5.65E-17	2	74.82	(255, 297, 338)
$\Delta(\delta/\gamma)$	1.16	1.41	1.24	4.46E-09	2	38.46	(255, 297, 338)
$\Delta(\delta/\theta)$	3.21	3.63	3.24	9.97E-13	2	55.27	(255, 297, 338)
$\Delta(\theta/\alpha)$	-1.90	-5.28	-1.92	4.46E-09	2	38.46	(255, 297, 338)
$\Delta(\theta/\beta)$	-0.69	-0.26	-0.08	0.27665	2	2.57	(255, 297, 338)
$\Delta(\theta/\gamma)$	0.08	0.35	0.23	3.68E-22	2	98.71	(255, 297, 338)
$\Delta(\theta/\delta)$	1.03	1.42	0.97	5.59E-08	2	33.40	(255, 297, 338)
$\Delta(\theta/(\alpha+\beta))$	0.01	0.00	0.05	4.04E-05	2	20.23	(255, 297, 338)
$\Delta(\beta/(\alpha+\theta))$	-0.03	-0.11	-0.04	1.79E-18	2	81.73	(255, 297, 338)

Table D.2: Frequency ratio Analysis - SubjectID 3

Feature name	**Mean Frequency Ratio**			**P Value**	**df**	**Chi-Squ**	**No. of samples**
	Easy	Medium	Difficult				
$\Delta(\alpha/\beta)$	-0.34	-0.08	0.00	6.25E-55	2	249.62	(105, 205, 263)
$\Delta(\alpha/\gamma)$	-0.54	-0.43	-0.11	9.54E-11	2	46.15	(105, 205, 263)
$\Delta(\alpha/\delta)$	0.31	-0.27	-1.11	5.71E-12	2	51.78	(105, 205, 263)
$\Delta(\alpha/\theta)$	0.13	0.10	-0.34	5.87E-31	2	139.22	(105, 205, 263)
$\Delta(\beta/\alpha)$	2.28	0.89	-0.77	6.25E-55	2	249.62	(105, 205, 263)
$\Delta(\beta/\gamma)$	0.43	-0.82	-0.39	2.25E-12	2	53.64	(105, 205, 263)
$\Delta(\beta/\delta)$	1.72	0.93	-1.68	3.59E-73	2	333.62	(105, 205, 263)
$\Delta(\beta/\theta)$	2.04	-0.49	-4.51	2.14E-40	2	182.69	(105, 205, 263)
$\Delta(\gamma/\alpha)$	0.79	-0.30	-1.11	6.85E-32	2	143.52	(105, 205, 263)
$\Delta(\gamma/\beta)$	0.69	0.38	-0.35	9.54E-11	2	46.15	(105, 205, 263)
$\Delta(\gamma/\delta)$	-0.16	-0.01	0.01	2.25E-12	2	53.64	(105, 205, 263)
$\Delta(\gamma/\theta)$	0.57	0.42	-0.67	2.33E-48	2	219.35	(105, 205, 263)
$\Delta(\delta/\alpha)$	-4.23	-16.64	-0.42	5.71E-12	2	51.78	(105, 205, 263)
$\Delta(\delta/\beta)$	-3.97	-4.41	-0.02	2.14E-40	2	182.69	(105, 205, 263)
$\Delta(\delta/\gamma)$	-1.42	-1.53	-0.51	0.00024749	2	16.61	(105, 205, 263)
$\Delta(\delta/\theta)$	-8.09	-9.18	-0.04	6.85E-32	2	143.52	(105, 205, 263)
$\Delta(\theta/\alpha)$	0.26	-0.27	-0.59	0.00024749	2	16.61	(105, 205, 263)
$\Delta(\theta/\beta)$	-0.78	-1.82	0.21	5.87E-31	2	139.22	(105, 205, 263)
$\Delta(\theta/\gamma)$	-0.99	-0.63	0.09	3.59E-73	2	333.62	(105, 205, 263)
$\Delta(\theta/\delta)$	-1.83	-1.71	0.11	2.33E-48	2	219.35	(105, 205, 263)
$\Delta(\theta/(\alpha+\beta))$	-0.01	-0.02	0.15	1.18E-42	2	193.08	(105, 205, 263)
$\Delta(\beta/(\alpha+\theta))$	0.24	0.13	0.07	2.71E-54	2	246.69	(105, 205, 263)

Table D.3: Frequency ratio Analysis - SubjectID 4

Feature name	**Mean Frequency Ratio**			**P Value**	**df**	**Chi-Squ**	**No. of samples**
	Easy	Medium	Difficult				
$\Delta(\alpha/\beta)$	-0.07	-0.20	-0.01	1.41E-49	2	224.97	(277, 365, 539)
$\Delta(\alpha/\gamma)$	-0.04	-0.22	0.12	8.13E-74	2	336.59	(277, 365, 539)
$\Delta(\alpha/\delta)$	-0.48	-1.85	-1.90	1.37E-34	2	155.95	(277, 365, 539)
$\Delta(\alpha/\theta)$	-1.12	-0.90	-0.50	3.30E-24	2	108.13	(277, 365, 539)
$\Delta(\beta/\alpha)$	-0.52	1.16	-0.39	1.41E-49	2	224.97	(277, 365, 539)
$\Delta(\beta/\gamma)$	-0.08	0.25	0.58	1.43E-30	2	137.44	(277, 365, 539)
$\Delta(\beta/\delta)$	-6.32	-1.57	-2.02	0.12121	2	4.22	(277, 365, 539)
$\Delta(\beta/\theta)$	-2.03	-4.18	-7.91	2.33E-28	2	127.25	(277, 365, 539)
$\Delta(\gamma/\alpha)$	-1.65	-3.79	-9.05	7.33E-52	2	235.49	(277, 365, 539)
$\Delta(\gamma/\beta)$	-0.70	0.23	-1.94	8.13E-74	2	336.59	(277, 365, 539)
$\Delta(\gamma/\delta)$	-0.19	-0.18	-0.39	1.43E-30	2	137.44	(277, 365, 539)
$\Delta(\gamma/\theta)$	-2.59	-1.37	-2.53	0.00011612	2	18.12	(277, 365, 539)
$\Delta(\delta/\alpha)$	1.09	1.84	2.07	1.37E-34	2	155.95	(277, 365, 539)
$\Delta(\delta/\beta)$	0.11	0.26	0.51	2.33E-28	2	127.25	(277, 365, 539)
$\Delta(\delta/\gamma)$	-1.15	-0.43	0.30	7.97E-43	2	193.87	(277, 365, 539)
$\Delta(\delta/\theta)$	0.42	0.49	0.99	7.33E-52	2	235.49	(277, 365, 539)
$\Delta(\theta/\alpha)$	-0.35	-1.05	-2.36	7.97E-43	2	193.87	(277, 365, 539)
$\Delta(\theta/\beta)$	1.48	2.16	0.75	3.30E-24	2	108.13	(277, 365, 539)
$\Delta(\theta/\gamma)$	0.25	0.39	0.26	0.12121	2	4.22	(277, 365, 539)
$\Delta(\theta/\delta)$	0.62	0.77	0.72	0.00011612	2	18.12	(277, 365, 539)
$\Delta(\theta/(\alpha+\beta))$	0.19	0.18	0.07	2.60E-28	2	127.03	(277, 365, 539)
$\Delta(\beta/(\alpha+\theta))$	0.03	0.11	0.04	3.03E-37	2	168.18	(277, 365, 539)

Table D.4: Frequency ratio Analysis - SubjectID 5

Feature name	**Mean Frequency Ratio**			**P Value**	**DOF**	**Chi-Squ**	**No. of samples**
	Easy	Medium	Difficult				
$\Delta(\alpha/\beta)$	-0.35	-0.03	-0.17	1.58E-51	2	233.95	(113, 224, 291)
$\Delta(\alpha/\gamma)$	-1.12	-0.90	-1.01	4.74E-20	2	88.99	(113, 224, 291)
$\Delta(\alpha/\delta)$	0.33	-0.69	0.31	9.24E-06	2	23.18	(113, 224, 291)
$\Delta(\alpha/\theta)$	-0.22	-0.29	-0.75	6.88E-21	2	92.85	(113, 224, 291)
$\Delta(\beta/\alpha)$	2.54	-0.98	1.66	1.58E-51	2	233.95	(113, 224, 291)
$\Delta(\beta/\gamma)$	-0.90	-1.80	-2.09	1.63E-06	2	26.65	(113, 224, 291)
$\Delta(\beta/\delta)$	0.92	-0.98	-0.76	1.79E-33	2	150.81	(113, 224, 291)
$\Delta(\beta/\theta)$	3.16	-0.60	2.87	3.77E-42	2	190.76	(113, 224, 291)
$\Delta(\gamma/\alpha)$	2.23	0.54	2.00	1.15E-32	2	147.09	(113, 224, 291)
$\Delta(\gamma/\beta)$	1.93	-0.64	1.70	4.74E-20	2	88.99	(113, 224, 291)
$\Delta(\gamma/\delta)$	0.20	0.09	0.22	1.63E-06	2	26.65	(113, 224, 291)
$\Delta(\gamma/\theta)$	0.87	-0.37	0.13	5.68E-19	2	84.03	(113, 224, 291)
$\Delta(\delta/\alpha)$	-0.81	-1.32	-2.18	9.24E-06	2	23.18	(113, 224, 291)
$\Delta(\delta/\beta)$	-0.86	-0.17	-0.85	3.77E-42	2	190.76	(113, 224, 291)
$\Delta(\delta/\gamma)$	-0.68	-0.88	-2.65	1.39E-39	2	178.95	(113, 224, 291)
$\Delta(\delta/\theta)$	-2.34	-0.56	-4.63	1.15E-32	2	147.09	(113, 224, 291)
$\Delta(\theta/\alpha)$	0.94	0.14	1.31	1.39E-39	2	178.95	(113, 224, 291)
$\Delta(\theta/\beta)$	0.34	-0.20	0.82	6.88E-21	2	92.85	(113, 224, 291)
$\Delta(\theta/\gamma)$	-0.47	0.05	0.06	1.79E-33	2	150.81	(113, 224, 291)
$\Delta(\theta/\delta)$	-1.54	-0.47	-0.78	5.68E-19	2	84.03	(113, 224, 291)
$\Delta(\theta/(\alpha+\beta))$	0.12	0.05	0.18	2.84E-19	2	85.41	(113, 224, 291)
$\Delta(\beta/(\alpha+\theta))$	0.19	0.02	0.11	5.80E-37	2	166.88	(113, 224, 291)

Table D.5: Frequency ratio Analysis - SubjectID 6

Feature name	**Mean Frequency Ratio**			**P Value**	**DOF**	**Chi-Squ**	**No. of samples**
	Easy	Medium	Difficult				
$\Delta(\alpha/\beta)$	0.16	0.69	0.79	2.67E-68	2	311.18	(106, 236, 629)
$\Delta(\alpha/\gamma)$	1.37	2.33	2.90	6.33E-81	2	369.33	(106, 236, 629)
$\Delta(\alpha/\delta)$	0.03	-0.14	-0.81	8.14E-64	2	290.54	(106, 236, 629)
$\Delta(\alpha/\theta)$	-0.13	-0.63	-0.64	4.13E-19	2	84.66	(106, 236, 629)
$\Delta(\beta/\alpha)$	-0.45	-5.20	-4.81	2.67E-68	2	311.18	(106, 236, 629)
$\Delta(\beta/\gamma)$	0.97	0.63	1.92	4.69E-100	2	457.42	(106, 236, 629)
$\Delta(\beta/\delta)$	-0.30	-2.68	-4.81	1.44E-68	2	312.42	(106, 236, 629)
$\Delta(\beta/\theta)$	0.01	-1.61	-4.88	8.62E-84	2	382.53	(106, 236, 629)
$\Delta(\gamma/\alpha)$	-0.01	-0.77	-4.55	1.26E-97	2	446.24	(106, 236, 629)
$\Delta(\gamma/\beta)$	-0.38	-2.53	-4.89	6.33E-81	2	369.33	(106, 236, 629)
$\Delta(\gamma/\delta)$	-0.19	-0.13	-0.55	4.69E-100	2	457.42	(106, 236, 629)
$\Delta(\gamma/\theta)$	-0.20	-1.21	-4.90	2.12E-90	2	412.96	(106, 236, 629)
$\Delta(\delta/\alpha)$	-6.40	5.70	6.44	8.14E-64	2	290.54	(106, 236, 629)
$\Delta(\delta/\beta)$	-2.95	8.89	9.68	8.62E-84	2	382.53	(106, 236, 629)
$\Delta(\delta/\gamma)$	-3.55	-1.61	0.59	4.26E-59	2	268.81	(106, 236, 629)
$\Delta(\delta/\theta)$	4.10	28.96	32.44	1.26E-97	2	446.24	(106, 236, 629)
$\Delta(\theta/\alpha)$	0.20	-0.24	-0.59	4.26E-59	2	268.81	(106, 236, 629)
$\Delta(\theta/\beta)$	1.12	1.56	2.18	4.13E-19	2	84.66	(106, 236, 629)
$\Delta(\theta/\gamma)$	1.52	3.46	3.59	1.44E-68	2	312.42	(106, 236, 629)
$\Delta(\theta/\delta)$	6.99	11.26	12.12	2.12E-90	2	412.96	(106, 236, 629)
$\Delta(\theta/(\alpha+\beta))$	0.16	0.32	0.33	7.56E-22	2	97.27	(106, 236, 629)
$\Delta(\beta/(\alpha+\theta))$	-0.08	-0.29	-0.36	1.72E-79	2	362.72	(106, 236, 629)

Table D.6: Frequency ratio Analysis - SubjectID 8

Feature name	**Mean Frequency Ratio**			**P Value**	**df**	**Chi-Squ**	**No. of samples**
	Easy	Medium	Difficult				
$\Delta(\alpha/\beta)$	1.33	1.50	1.60	1.95E-33	2	150.63	(93, 186, 225)
$\Delta(\alpha/\gamma)$	1.09	2.39	3.00	9.02E-45	2	202.83	(93, 186, 225)
$\Delta(\alpha/\delta)$	1.99	1.04	1.52	2.06E-30	2	136.71	(93, 186, 225)
$\Delta(\alpha/\theta)$	2.68	2.31	2.65	2.08E-29	2	132.08	(93, 186, 225)
$\Delta(\beta/\alpha)$	-1.80	-3.26	-6.92	1.95E-33	2	150.63	(93, 186, 225)
$\Delta(\beta/\gamma)$	-2.93	-1.76	-0.17	6.17E-27	2	120.70	(93, 186, 225)
$\Delta(\beta/\delta)$	0.71	-1.12	-1.10	7.03E-34	2	152.67	(93, 186, 225)
$\Delta(\beta/\theta)$	0.71	-2.66	-2.78	9.53E-36	2	161.28	(93, 186, 225)
$\Delta(\gamma/\alpha)$	0.52	-0.85	-1.83	7.18E-44	2	198.68	(93, 186, 225)
$\Delta(\gamma/\beta)$	-0.22	-1.77	-5.03	9.02E-45	2	202.83	(93, 186, 225)
$\Delta(\gamma/\delta)$	0.30	0.02	-0.08	6.17E-27	2	120.70	(93, 186, 225)
$\Delta(\gamma/\theta)$	0.68	-0.29	-0.85	3.00E-50	2	228.06	(93, 186, 225)
$\Delta(\delta/\alpha)$	-4.48	-0.94	-3.28	2.06E-30	2	136.71	(93, 186, 225)
$\Delta(\delta/\beta)$	-1.44	0.44	-0.15	9.53E-36	2	161.28	(93, 186, 225)
$\Delta(\delta/\gamma)$	-0.69	0.43	0.00	6.80E-17	2	74.45	(93, 186, 225)
$\Delta(\delta/\theta)$	-10.12	0.58	-0.42	7.18E-44	2	198.68	(93, 186, 225)
$\Delta(\theta/\alpha)$	0.17	-0.80	-0.83	6.80E-17	2	74.45	(93, 186, 225)
$\Delta(\theta/\beta)$	-1.99	-1.16	-2.42	2.08E-29	2	132.08	(93, 186, 225)
$\Delta(\theta/\gamma)$	-0.52	0.15	0.07	7.03E-34	2	152.67	(93, 186, 225)
$\Delta(\theta/\delta)$	-4.26	-0.09	0.03	3.00E-50	2	228.06	(93, 186, 225)
$\Delta(\theta/(\alpha+\beta))$	-0.40	-0.24	-0.36	2.59E-31	2	140.86	(93, 186, 225)
$\Delta(\beta/(\alpha+\theta))$	-0.33	-0.37	-0.48	8.61E-43	2	193.72	(93, 186, 225)

Table D.7: Frequency ratio Analysis - SubjectID 9

Feature name	**Mean Frequency Ratio**			**P Value**	**df**	**Chi-Squ**	**No. of samples**
	Easy	Medium	Difficult				
$\Delta(\alpha/\beta)$	-0.50	-0.11	-0.29	1.55E-50	2	229.38	(280, 460, 521)
$\Delta(\alpha/\gamma)$	-0.03	0.79	0.25	5.97E-91	2	415.50	(280, 460, 521)
$\Delta(\alpha/\delta)$	-0.90	-1.70	-0.51	9.86E-88	2	400.68	(280, 460, 521)
$\Delta(\alpha/\theta)$	-0.87	-0.83	-0.63	4.70E-16	2	70.59	(280, 460, 521)
$\Delta(\beta/\alpha)$	2.93	-1.89	1.05	1.55E-50	2	229.38	(280, 460, 521)
$\Delta(\beta/\gamma)$	3.86	4.16	3.72	2.47E-47	2	214.63	(280, 460, 521)
$\Delta(\beta/\delta)$	-1.29	-3.99	-1.88	6.27E-49	2	221.98	(280, 460, 521)
$\Delta(\beta/\theta)$	-1.21	-7.32	-1.21	2.35E-165	2	758.14	(280, 460, 521)
$\Delta(\gamma/\alpha)$	-0.92	-5.59	-0.86	3.49E-175	2	803.41	(280, 460, 521)
$\Delta(\gamma/\beta)$	-0.58	-3.78	-1.65	5.97E-91	2	415.50	(280, 460, 521)
$\Delta(\gamma/\delta)$	-0.54	-0.69	-0.43	2.47E-47	2	214.63	(280, 460, 521)
$\Delta(\gamma/\theta)$	-1.30	-3.09	-1.47	1.65E-81	2	372.01	(280, 460, 521)
$\Delta(\delta/\alpha)$	24.81	25.24	23.52	9.86E-88	2	400.68	(280, 460, 521)
$\Delta(\delta/\beta)$	4.62	5.27	4.47	2.35E-165	2	758.14	(280, 460, 521)
$\Delta(\delta/\gamma)$	1.56	2.34	1.28	1.49E-42	2	192.62	(280, 460, 521)
$\Delta(\delta/\theta)$	28.76	29.83	28.28	3.49E-175	2	803.41	(280, 460, 521)
$\Delta(\theta/\alpha)$	-0.83	-1.88	-0.67	1.49E-42	2	192.62	(280, 460, 521)
$\Delta(\theta/\beta)$	7.33	6.94	6.62	4.70E-16	2	70.59	(280, 460, 521)
$\Delta(\theta/\gamma)$	0.87	1.39	0.83	6.27E-49	2	221.98	(280, 460, 521)
$\Delta(\theta/\delta)$	8.28	9.32	8.06	1.65E-81	2	372.01	(280, 460, 521)
$\Delta(\theta/(\alpha+\beta))$	0.34	0.40	0.25	7.24E-58	2	263.14	(280, 460, 521)
$\Delta(\beta/(\alpha+\theta))$	0.29	0.18	0.20	2.60E-12	2	53.35	(280, 460, 521)

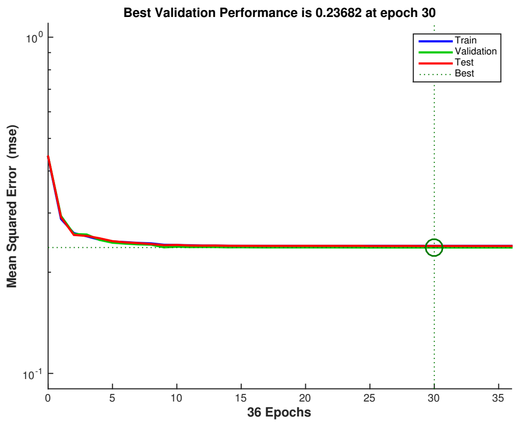
Table D.8: Frequency ratio Analysis - SubjectID 10

Feature name	**Mean Frequency Ratio**			**P Value**	**df**	**Chi-Squ**	**No. of samples**
	Easy	Medium	Difficult				
$\Delta(\alpha/\beta)$	0.84	0.77	0.94	5.75E-22	2	97.82	(264, 699, 429)
$\Delta(\alpha/\gamma)$	1.97	2.96	3.11	1.62E-57	2	261.53	(264, 699, 429)
$\Delta(\alpha/\delta)$	0.04	-0.99	-0.37	6.58E-153	2	700.82	(264, 699, 429)
$\Delta(\alpha/\theta)$	-0.53	-0.62	-0.38	9.28E-32	2	142.91	(264, 699, 429)
$\Delta(\beta/\alpha)$	-1.82	-1.94	-3.11	5.75E-22	2	97.82	(264, 699, 429)
$\Delta(\beta/\gamma)$	-1.26	0.68	1.07	1.46E-55	2	252.53	(264, 699, 429)
$\Delta(\beta/\delta)$	-1.47	-1.58	-1.92	0.00058827	2	14.88	(264, 699, 429)
$\Delta(\beta/\theta)$	-0.29	-3.51	-2.04	3.90E-121	2	554.50	(264, 699, 429)
$\Delta(\gamma/\alpha)$	-0.04	-1.46	-1.73	3.85E-140	2	642.03	(264, 699, 429)
$\Delta(\gamma/\beta)$	-0.48	-0.97	-2.72	1.62E-57	2	261.53	(264, 699, 429)
$\Delta(\gamma/\delta)$	0.05	-0.29	-0.34	1.46E-55	2	252.53	(264, 699, 429)
$\Delta(\gamma/\theta)$	-0.34	-0.92	-1.65	1.62E-65	2	298.37	(264, 699, 429)
$\Delta(\delta/\alpha)$	-3.71	2.76	2.05	6.58E-153	2	700.82	(264, 699, 429)
$\Delta(\delta/\beta)$	2.19	4.28	4.35	3.90E-121	2	554.50	(264, 699, 429)
$\Delta(\delta/\gamma)$	-3.10	-0.30	-0.30	1.45E-139	2	639.38	(264, 699, 429)
$\Delta(\delta/\theta)$	5.21	14.03	13.74	3.85E-140	2	642.03	(264, 699, 429)
$\Delta(\theta/\alpha)$	1.42	-0.35	0.50	1.45E-139	2	639.38	(264, 699, 429)
$\Delta(\theta/\beta)$	4.78	5.11	4.51	9.28E-32	2	142.91	(264, 699, 429)
$\Delta(\theta/\gamma)$	7.68	7.78	7.71	0.00058827	2	14.88	(264, 699, 429)
$\Delta(\theta/\delta)$	23.04	24.39	24.05	1.62E-65	2	298.37	(264, 699, 429)
$\Delta(\theta/(\alpha+\beta))$	0.28	0.40	0.27	1.87E-75	2	344.13	(264, 699, 429)
$\Delta(\beta/(\alpha+\theta))$	-0.24	-0.19	-0.30	6.86E-25	2	111.28	(264, 699, 429)

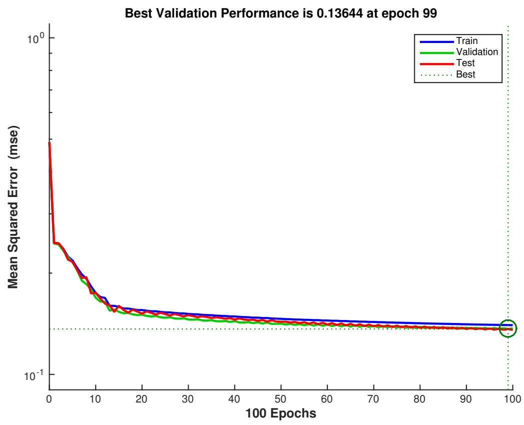
APPENDIX E

Classification plots

This chapter provides the performance and regression plots for Method 1 and Method 2. (Refer section 6.4 for more information on different methods used for classification).

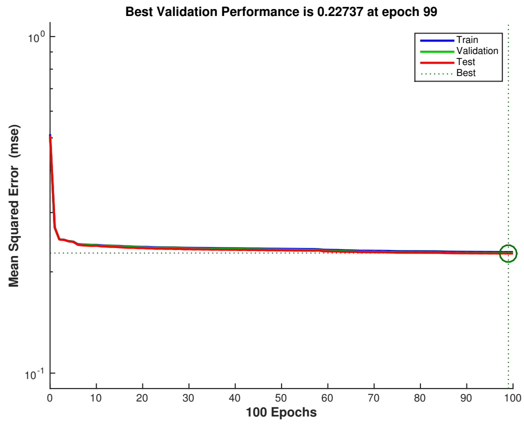


(a) Performance plot (easy vs medium vs difficult)

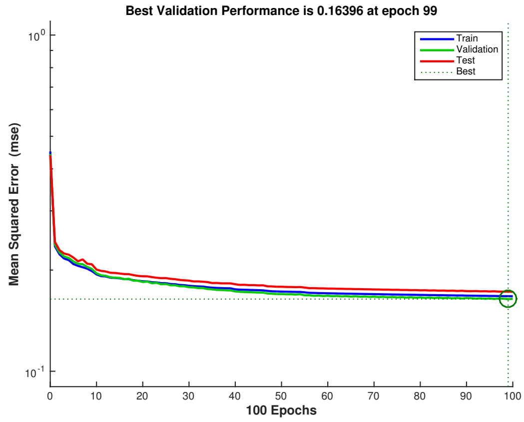


(b) Performance plot (easy vs difficult)

Figure E.1: Method1 - Classification performance plot



(a) Performance plot (easy vs medium vs difficult)



(b) Performance plot (easy vs difficult)

Figure E.2: Method2 - Classification performance plot

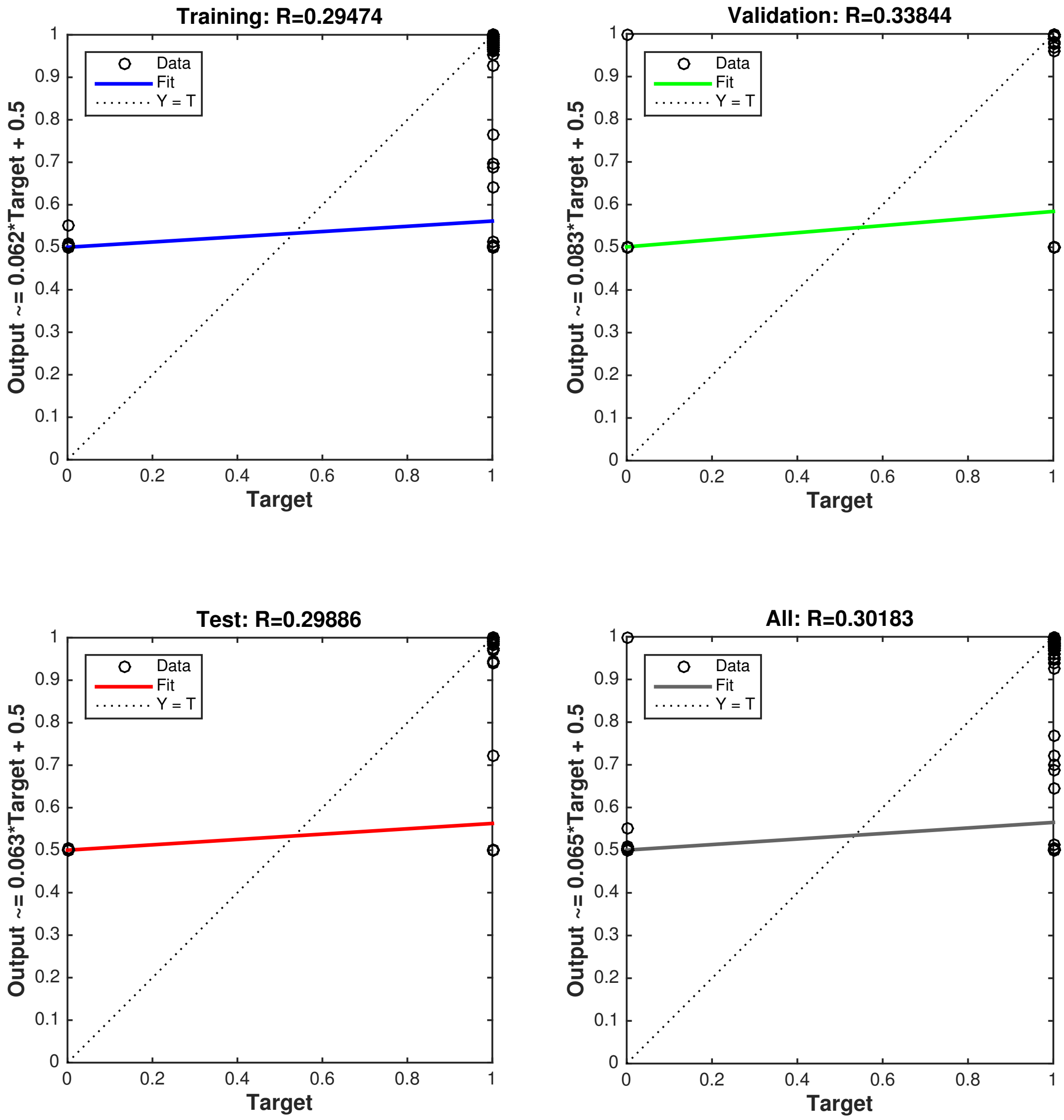


Figure E.3: Method 1 - Regression Plot (easy vs medium vs difficult)

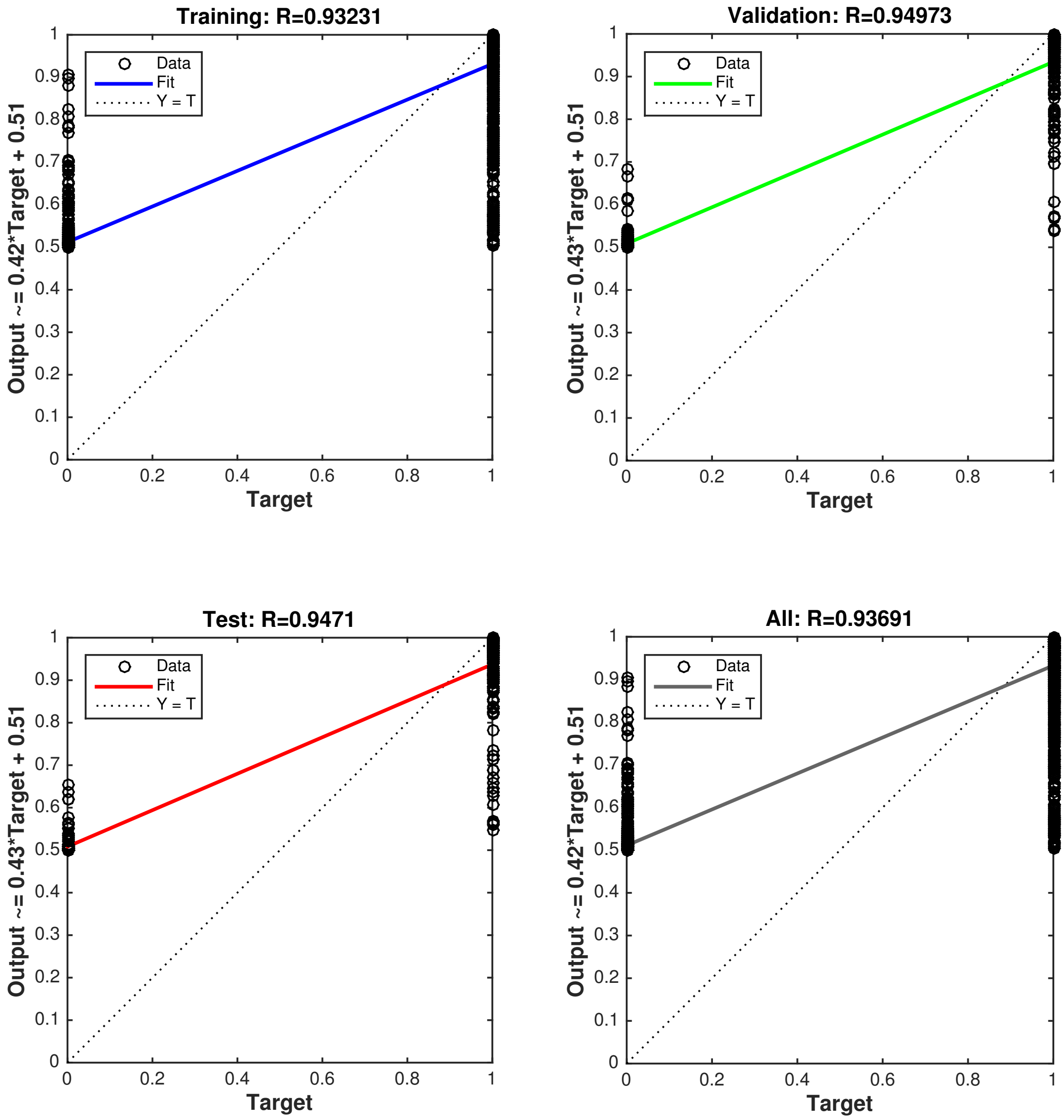


Figure E.4: Method 1 - Regression Plot (easy vs difficult)

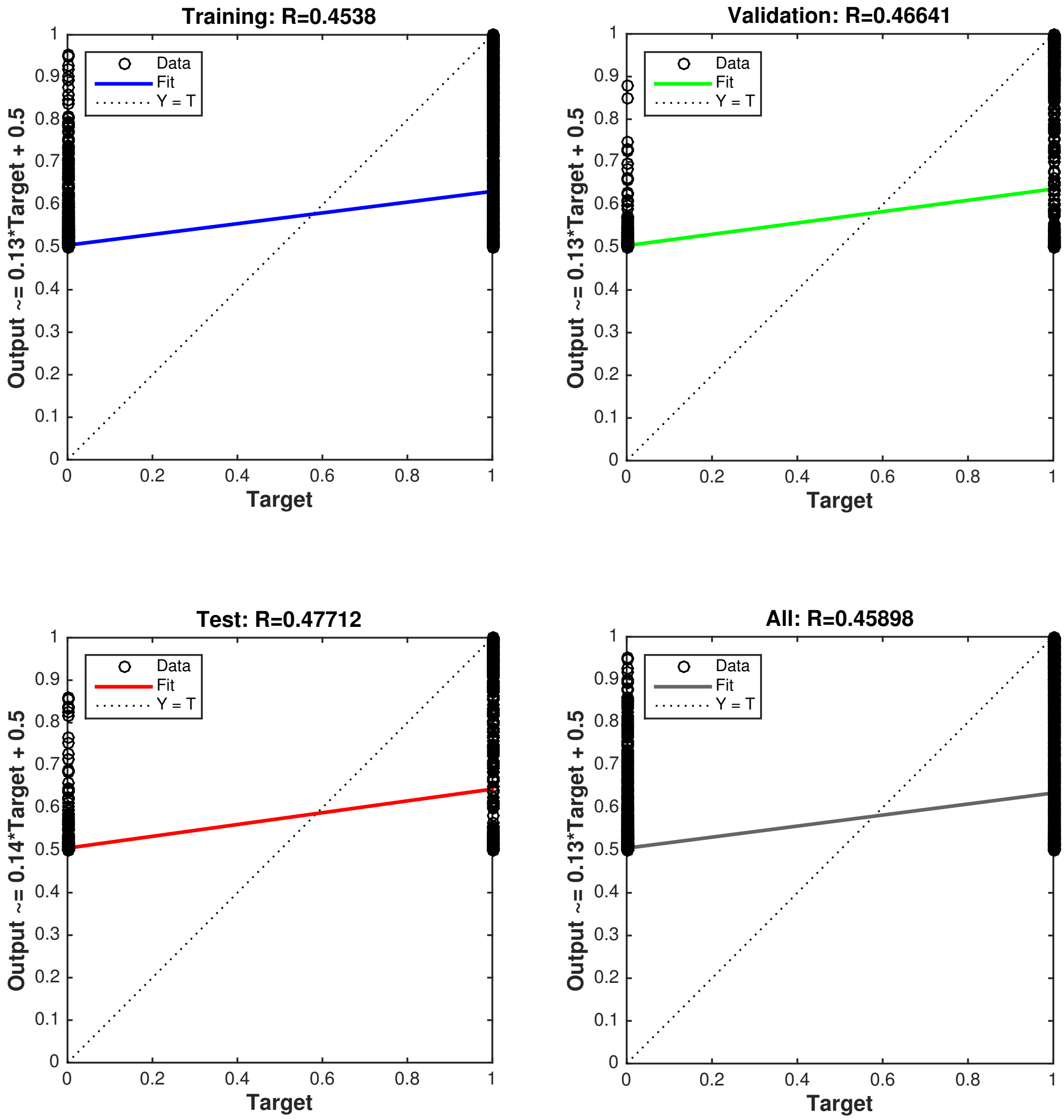


Figure E.5: Method 2 - Regression Plot (easy vs medium vs difficult)

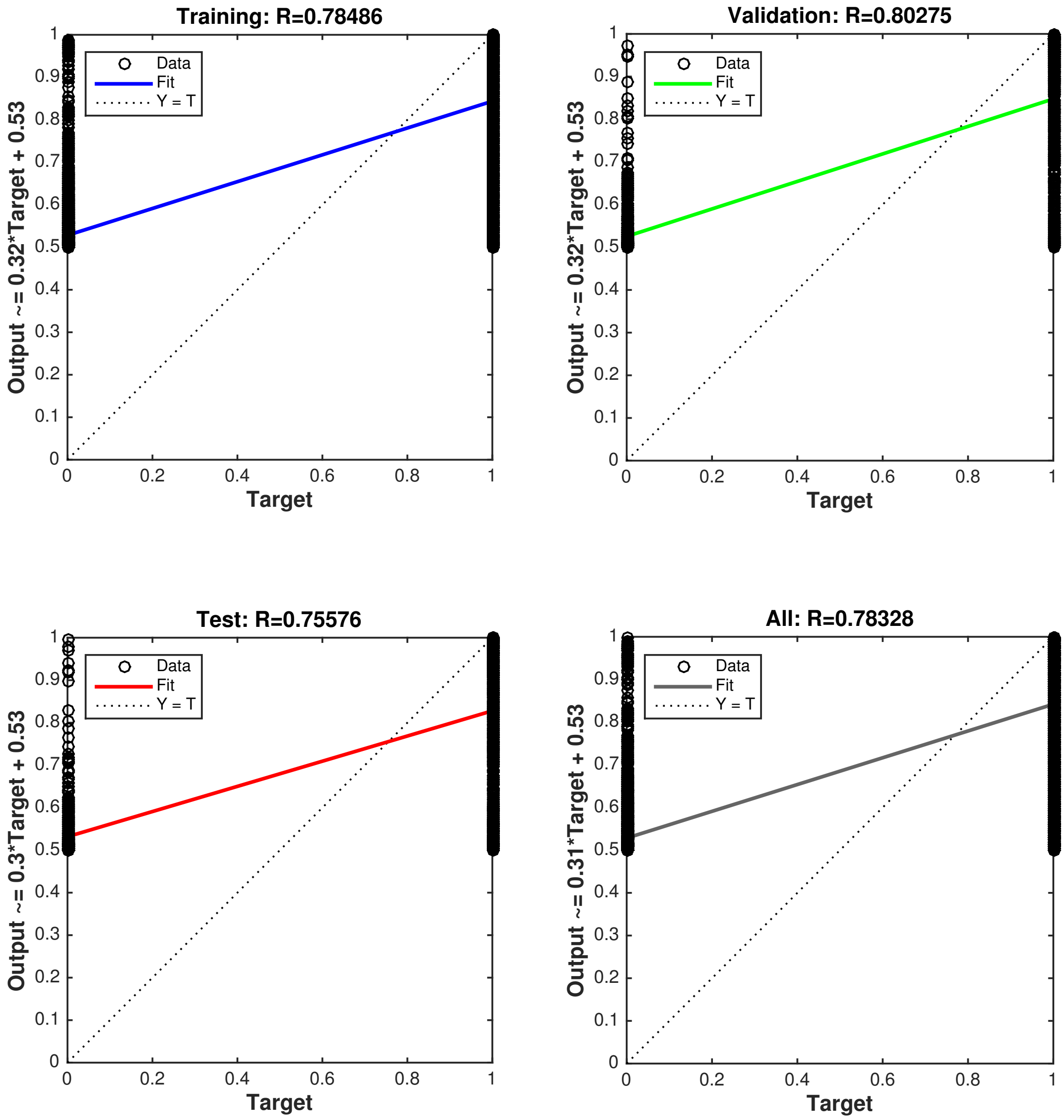


Figure E.6: Method 2 - Regression Plot (easy vs difficult)

APPENDIX F

Task Presentation System User Guide

This document covers the workflow of the GUI and explains each screen in detail. The Graphical User Interface (GUI) used in this research is written in Matlab, as it is easier to collect the participant solution data and intercept with the Emotive EEG device via COM port. This user guide document is intended for the participants who volunteered for this research experiment. This document includes the functionality of how the GUI works and describes each front-end screen in following sections.

GUI Workflow

This GUI consists of following front-end screens.

- Welcome Screen
- Instructions Screen
- Program Task Screen
- Relax Screen

- Thank you Screen

The workflow of the GUI is defined as when the participant is ready, GUI program in Matlab is executed. The first screen would be Welcome screen which gives general information about the project thesis. When the next button is clicked, the instruction screen is displayed that shows the instructions to follow during the experiment. The user reads the instructions and enters their name in the text box provided and clicks next. This name will be used as an identification of the experimental data saved after completion of the experiment. Then the Relax screen will displayed, where the user use this time to relax. This relax screen will be displayed in-between each tasks and before the start of the actual experiment. There is no time restrictions for the user to relax, so the relax screen can be skipped by the user at any time.

Once the Relax screen is skipped it displays the program tasks where the Java program, which is to be solved, will be displayed in the screen. There are no time restrictions for the user to solve the tasks. Once the program is solved, the solution to the problem will be entered in the text box provided in the screen and clicked next. Then a relax screen will be displayed before the user moves to the next question. This way the task screen and the relax screen is displayed orderly till all the Java programs are solved. Once completed the thank you screen is displayed.

Front-end Screens

This section explains about the screens displayed to the user and its functionalities.

Welcome Screen

Welcome screen (Fig. F.1) is the one displayed first to the user when the program is executed. It displays information about the Project title, researcher details and the Supervisor details to the participants. When the next button is clicked it moves to the Instruction screen (Fig. F.2).

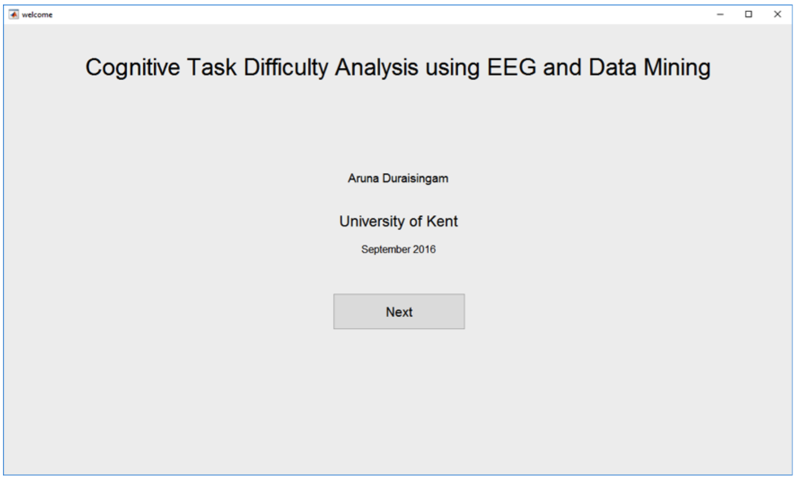


Figure F.1: Welcome Screen

Instruction Screen

Instruction screen displays the instructions to be followed by the participant during the experiment. This screen consists of the instructions to follow during the experiment and a textbox to enter the participant name. Once the instructions are read, user enters their name in the text box provided in the screen. When the user tries to start the experiment by clicking the start button without entering name, it displays an error message âĂIJName field cannot be emptyâĂİ. The entered name is used to identify the person and their corresponding data while saving. When the start button is clicked it moves to the relax screen (Fig. F.3).

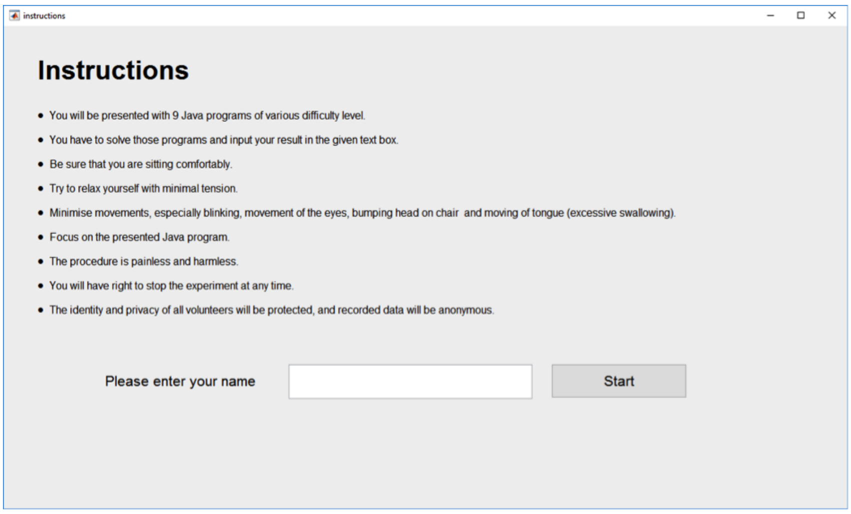


Figure F.2: Instruction Screen

Relax Screen

Relax screen is displayed for the user to relax before starting to solve the Java tasks or in-between each Java tasks. This relax time can be used to relax or fill the NASA TLX questionnaire (after solving the program tasks). There is no time restrictions for the user to Relax, the user can decide to move to next Program Task (when the program tasks are displayed) at any time by clicking the skip button in the screen. There is time counter processed in background of the relax screen for a minute. Once the 1 minute time counter is over it displays âĂIJReady for next questionâĂİ, to intimate the user that the person ahs relaxed for a minute. Once the relax screen is clicked the Program tasks screen will be displayed (Fig. F.4).

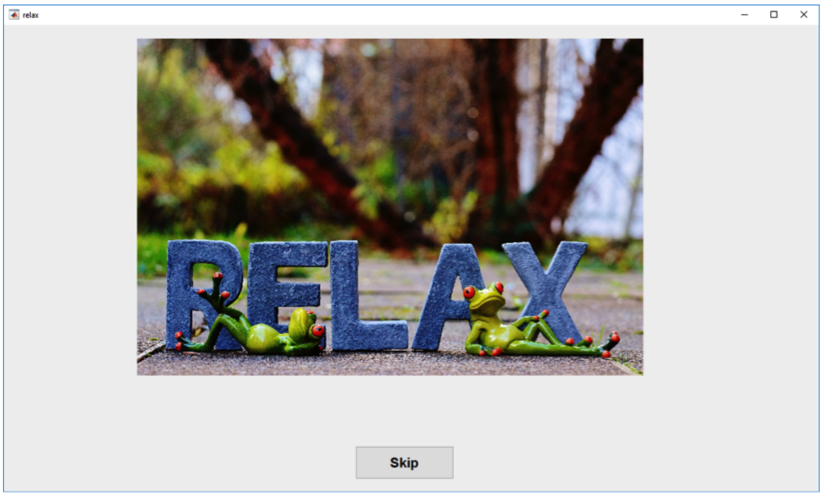


Figure F.3: Relax Screen

Program Task Screen

This screen displays the experimental java programs the user has to solve. The left hand side will be the program syntax and the right hand side contains the text box for the user to enter the solution of the program. Once the solution is entered and clicked next button the relax screen will be displayed for the user to relax. This process of displaying task screen and relax screen will happen till all the java programs are solved. Once all the programs are solved, the thank you screen will be displayed (Fig. F.5).

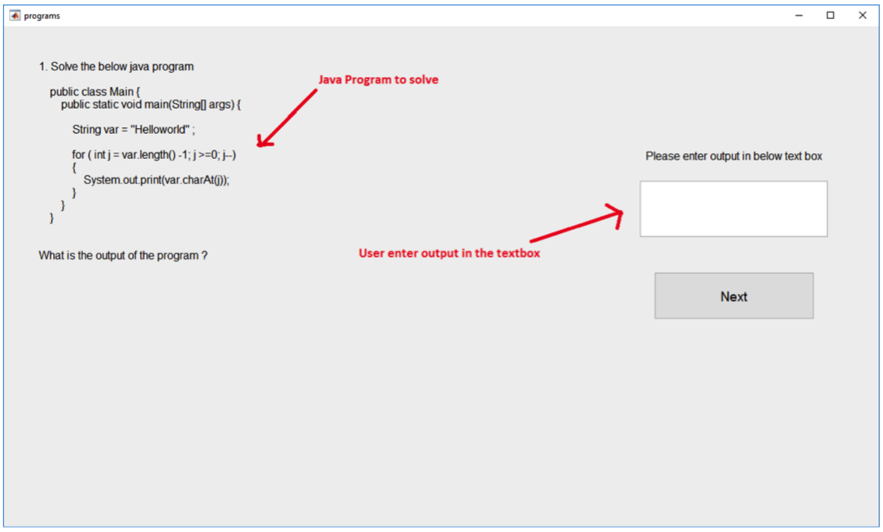


Figure F.4: Task Screen

Thank you screen

This screen defines the end of the experiment. During this screen the user solution data will be saved.

Figure F.5: Thank you Screen

APPENDIX G

Experimental Program Comprehension Tasks

1. Solve the below java program

```
public class Main {
    public static void main(String[] args) {

        String var = "Helloworld" ;

        for ( int j = var.length() -1; j >=0; j--)
        {
            System.out.print(var.charAt(j));
        }
    }
}
```

What is the output of the program?

2. Solve the below java program

```
public class Main {
    public static void main(String[] args) {

    int [] vars = {4, 8, 10, 12, 16, 10, 18, 2, 3, 5};
    int value = 0;
    int count = 0;

    for(int i = 0; i< vars.length; i++){
        value += vars[i];
        count++;
    }

    int temp = value/count;
    System.out.println(temp);
    }
}
```

What is the output of the program?

___

3. Solve the below java program

```
public class Main {
    public static void main(String[] args) {
```

```
        double [] array = {10, 15, 5, 7, 9, 4};
        double service_tax_percent = 5;
        double tip_percent = 10;
        double total_amount = 0;
        double service_tax_amount = 0;
        double tip_amount = 0;

        for(int i=0; i< array.length; i++){
            total_amount = total_amount + array[i];
        }

        service_tax_amount =
            (total_amount * service_tax_percent)/100.00 ;

        total_amount = total_amount + service_tax_amount;
        tip_amount = (total_amount * tip_percent) / 100.00;
        System.out.println("Tip amount is : Âč"+ tip_amount);
        }
    }
```

What is the output of the program?

____

4. Solve the below java program

```
    public class Main {

        public static void main(String[] args) {

            ArrayList<Object> array =
```

```
                    new ArrayList<Object> ();

            String a = "Circle";
            String b = "Triangle";
            String c = "Square";
            String d = "Triangle";
            String e= "Circle";
            String f = "Triangle";

            array.add(b);
            array.add(a);
            array.add(d);
            array.add(e);

            for (int i = 1; i < 4; i++) {
                System.out.print(array.get(i) + " ");
            }

        }
    }
```

Last three shape objects are?

____

5. Solve the below java program

```
    public class Main {
        public static void main(String[] args) {

            Days d1 = Days.TH;
```

```java
        Days d2 = Days.M;

        for (Days d: Days.values()){

            if(d.equals(Days.F))
                break;

            d2 = d;
        }
        System.out.println((d1 == d2)
            ? "Same" : "Not Same");
    }

    enum Days{M, T, W, TH, F, SA, SU};
}
```

What is the output of the program?

____

6. Solve the below java program

```java
public class Main {
    public static void main(String[] args) {

        ArrayList<Object> array =
            new ArrayList<Object>();

        array.add("Triangle");
        array.add("Circle");
        array.add("Square");
```

```java
            array.add("Circle");
            array.add("Triangle");
            array.add("Square");
            array.add("Square");

            swap(array, 5, 4);
            swap(array, 5, 6);

            for (int i = 4; i < 7; i++) {
                System.out.print(array.get(i) + " ");
            }
        }

    public static void swap(ArrayList array, int i, int j){
        Object temp = array.get(j);
        array.set(j, array.get(i));
        array.set(i, temp);
    }
    }
```

Last three shapes are?

____

7. Solve the below java program
(Note: Rectangle object is defined as drawRectangle(x-axis, y-axis, width,

```java
    class MyCanvas extends JComponent {
        public void paint(Graphics g) {

            g.drawRect (10, 10, 10, 10);
```

```
            g.drawRect(30, 50, 10, 10);
        }
    }


Will two rectangles overlap?

_______________________________________________________
```

```
8. Solve the below java program.
    Please check on the order of the variable displayed.
(Note: Rectangle object is defined as
    drawRectangle(x-axis, y-axis, width, height))

    class MyCanvas extends JComponent {
        public void paint(Graphics g) {

            int x1 = 10;
            int y1 = 10;
            int x2 = 130;
            int y2 = 50;

            int w1 = 130;
            int h1 = 90;
            int w2 = 100;
            int h2 = 100;

            g.drawRect (x1, y1, w1, h1);
            g.drawRect(x2, y2, w2, h2);

        }
```

```
    }
```

```
Will the two rectangles overlap?
```

```
9. Solve the below java program
(Note: Rectangle object is defined as
        drawRectangle(x-axis, y-axis, width, height))

    class MyCanvas extends JComponent {
        public void paint(Graphics g) {

            int x1 = 10;
            int y1 = 10;
            int w1 = 15;
            int h1 = 15;

            x1 = x1 * 2;
            y1 = y1 * 2;

            int x2 = 15;
            int y2 = 20;
            int w2 = 15;
            int h2 = 10;

            x2 = x2 * 2;
            y2 = y2 * 2;

            int x3 = 10;
            int y3 = 45;
```

```
            int w3 = 25;
            int h3 = 10;

            x3 = x3 * 2;
            y3 = y3 * 2;

            g.drawRect (x1, y1, w1, h1); // Rectangle1
            g.drawRect(x2, y2, w2, h2); // Rectangle2
            g.drawRect(x3, y3, w3, h3); // Rectangle3

        }
    }
```

Which two rectangles overlap?

APPENDIX H

Practise Program Comprehension Tasks

```
1. Solve the below java program

   public class Main {
       public static void main(String[] args) {

           int[] vars = {1, 2, 3, 4, 5};
           int value = 0;

           for(int i = 0; i< vars.length; i++){
               value += vars[i];
           }

           System.out.println(value);
       }
   }
```

What is the output of the program?

____

2. Solve the below java program

```
public class Main {
    public static void main(String[] args) {

        ArrayList<Object> array =
                new ArrayList<Object>();

        int temp1 = 21;

        array.add("Tringle");
        array.add("Square");
        array.add("Triangle");

        Object o = (17 >= temp1)? "Triangle" : "Square";

        array.add(o);

        System.out.println("Last Three Shape Object are:");

        for(int i=1; i< 4; i++){
            System.out.print(array.get(i) + " ");
        }
    }
}
```

What is the output of the program?

3. Solve the below java program
(Note: Rectangle object is defined as

```
    drawRectangle(x-axis, y-axis, width, height))

    class MyCanvas extends JComponent {
        public void paint(Graphics g) {

            g.drawRect (2, 2, 10, 10);
            g.drawRect(5, 5, 10, 10);

        }
    }
```

Will two rectangles overlap?

APPENDIX I

Questionnaire

Pre-Questionnaire

```
1. Solve the below java program and fill the output
   and analysis section.

public class Main {
    public static void main(String[] args) {
        int [] vars = {12, 14, 16, 10, 13, 9, 8, 10};
        int value = 0;

        for(int i = 0; i< vars.length; i++){
            value += vars[i];
        }

        System.out.println(value);
```

```
    }
}
```

What is the output of the program? ______________

Analysis

1. Time spent to solve the problem _______________

2. Mental workload rate_________________
(Rate from 1 to 10
Where, 1 – very less mental workload used
10 – very high mental workload used)

3. General Task Difficulty _________________
(Rate from 1 to 10
Where, 1 âĂŞ very easy task
10 âĂŞ Impossible to solve with only mind)

---

2. Solve the below java program and fill the output and analysis section.

```
public class Main {
    public static void main(String[] args) {
        int [] vars = {4, 5, 3, 10, 6, 9, 7};
        int value = 0;

          for(int i = 0; i< vars.length; i++){
              vars[i] = vars[i]*2
        }
```

```
        for(int i = 0; i< vars.length; i++){
            value += vars[i];
        }
        System.out.println(value);
    }
}
```

Output

What is the output of the program? _____________

Analysis

1. Time spent to solve the problem ______________

2. Mental workload rate_______________
(Rate from 1 to 10
    Where, 1 – very less mental workload used
    10 – very high mental workload used)

3. General Task Difficulty ________________
(Rate from 1 to 10
    Where, 1 âĂŞ very easy task
    10 âĂŞ Impossible to solve with only mind)

---

3. Solve the below java program and fill the output
    and analysis section.

```
public class Main {
```

```
    public static void main(String[] args) {

        int[] vars = {4, 8, 10, 12, 16, 10, 18, 2, 3, 5};
        int value = 0;
        int count = 0;

        for(int i = 0; i< vars.length; i++){
            value += vars[i];
            count++;
        }

        int temp = value/count;

        System.out.println(temp);

    }
}
```

Output

What is the output of the program? _____________

Analysis

1. Time spent to solve the problem ______________

2. Mental workload rate______________

(Rate from 1 to 10

Where, 1 – very less mental workload used

10 – very high mental workload used)

3. General Task Difficulty ________________

(Rate from 1 to 10
    Where, 1 âĂŞ very easy task
    10 âĂŞ Impossible to solve with only mind)

---

4. Solve the below java program and fill the output
    and analysis section.

```
public class Main {
      public static void main(String[] args) {
String var = "Helloworld" ;
for ( int j = var.length() -1; j >=0; j--)
{
System.out.print(var.charAt(j));
}
     }
}
```

Output

What is the output of the program? ____________

Analysis

1. Time spent to solve the problem _____________

2. Mental workload rate______________
(Rate from 1 to 10
    Where, 1 – very less mental workload used
    10 – very high mental workload used)

```
3.      General Task Difficulty  _______________
(Rate from 1 to 10
    Where, 1 âĂŞ very easy task
    10 âĂŞ Impossible to solve with only mind)
```

```
5. Solve the below java program and fill the output
    and analysis section.

public class Main {
    public static void main(String[] args) {

    double [] array  = {10, 15, 5, 7, 9, 4};
    double service_tax_percent = 5;
    double tip_percent = 10;
    double total_amount = 0,
    double service_tax_amount = 0;
    double tip_amount = 0;

    for(int i=0; i< array.length;  i++){
        total_amount = total_amount + array[i];
    }

    service_tax_amount =
            (total_amount * service_tax_percent)/100.00 ;

    total_amount = total_amount + service_tax_amount;
    tip_amount = (total_amount * tip_percent) / 100.00;
    System.out.println("Tip␣amount␣is␣:␣Âč"+ tip_amount);
    }
}
```

Output

What is the output of the program? _____________

Analysis

1. Time spent to solve the problem ______________

2. Mental workload rate______________
(Rate from 1 to 10
Where, 1 – very less mental workload used
10 – very high mental workload used)

3. General Task Difficulty ________________
(Rate from 1 to 10
Where, 1 âĂŞ very easy task
10 âĂŞ Impossible to solve with only mind)

6. Solve the below java program and fill the output and analysis section.

```
public class Main {
    public static void main(String[] args) {
        ArrayList<Object> array =
                    new ArrayList<Object> ();

        String a = "Circle";
        String b = "Triangle";
        String c = "Square";
        String d = "Triangle";
```

```
        String e= "Circle";
        String f = "Triangle";

        array.add(b);
        array.add(a);
        array.add(d);
        array.add(e);

        for (int i = 1; i < 4; i++) {
            System.out.print(array.get(i) + " ");
        }
    }
}
Output

Last three shape objects are ?    ____________

Analysis

1.        Time spent to solve the problem _____________

2.        Mental workload rate______________
(Rate from 1 to 10
    Where, 1 - very less mental workload used
    10 - very high mental workload used)

3.        General Task Difficulty   _______________
(Rate from 1 to 10
    Where, 1 âĂŞ very easy task
    10 âĂŞ Impossible to solve with only mind)
```

```
7. Solve the below java program and fill the output
   and analysis section.

public class Main {
    public static void main(String[] args) {
        ArrayList<Object> array =
                new ArrayList<Object>();

        array.add("Triangle");
        array.add("Circle");
        array.add("Square");
        array.add("Circle");
        array.add("Triangle");
        array.add("Square");
        array.add("Square");
        array.add("Circle");

        swap(array, 5, 4);
        swap(array, 5, 7);

        for (int i = 1; i < 4; i++) {
            System.out.print(array.get(i) + " ");
        }
    }

    public static void swap(ArrayList array, int i, int j){
        Object temp = array.get(j);
        array.set(j, array.get(i));
        array.set(i, temp);
```

```
    }
}
Last three shape objects are ?    ____________

Analysis

1.      Time spent to solve the problem _____________

2.      Mental workload rate______________
(Rate from 1 to 10
    Where, 1 − very less mental workload used
    10 − very high mental workload used)

3.      General Task Difficulty  _______________
(Rate from 1 to 10
    Where, 1 âĂŞ very easy task
    10 âĂŞ Impossible to solve with only mind)
```

```
8. Solve the below java program and fill the output
    and analysis section.

public class Main {
    public static void main(String[] args) {

        ArrayList<Object> array =
                new ArrayList<Object>();

        Object o = "Circle";
        array.add(o);
        o = "Square";
```

```java
        array.add(o);
        o = "Triangle";
        array.add(o);
        o = "Square";
        array.add(o);

        int i=1;
        int temp = 10;

        System.out.println("Last␣Three␣Shape␣Object␣are:");

        while (i<5){
            temp = temp % 3 ;
            System.out.print(array.get(temp) + "␣");
            temp = temp + 2;
            i++ ;
        }
    }
}
```

What is the output of the program? ____________

Analysis

1. Time spent to solve the problem ______________

2. Mental workload rate______________
(Rate from 1 to 10
Where, 1 – very less mental workload used
10 – very high mental workload used)

```
3.      General Task Difficulty  _______________
(Rate from 1 to 10
    Where, 1 âĂŞ very easy task
    10 âĂŞ Impossible to solve with only mind)
```

```
9. Solve the below java program and fill the output
    and analysis section.

public class Main {
    public static void main(String[] args) {
        ArrayList<Object> array =
                new ArrayList<Object>();

        int temp1 = 21;
        int temp2 = 11;

        array.add("Tringle");
        array.add("Square");
        array.add("Triangle");

        Object o = (17 >= temp1) ?
                ((temp2 > 17) ? "Triangle" : "Square") :
                ( (temp1 < temp2) ? "Circle" : "Square");

        array.add(o);

        System.out.println("Last␣Three␣Shape␣Object␣are:");

        for(int i=1; i< 4; i++){
            System.out.print(array.get(i) + "␣");
```

```
        }
    }
}
```

What is the output of the program? ____________

Analysis

1. Time spent to solve the problem _____________

2. Mental workload rate______________

(Rate from 1 to 10

Where, 1 – very less mental workload used

10 – very high mental workload used)

3. General Task Difficulty _______________

(Rate from 1 to 10

Where, 1 âĂŞ very easy task

10 âĂŞ Impossible to solve with only mind)

---

10. Solve the below java program and fill the output and analysis section.

```
public class Main {
      public static void main(String[] args) {

          Days d1 = Days.TH;
          Days d2 = Days.M;

           for (Days d: Days.values()){
                   if(d.equals(Days.F))
```

```
                        break;
                d2 = d;
        }

        System.out.println((d1 == d2)
                ? "Same" : "Not_Same");
    }

    enum Days{M, T, W, TH, F, SA, SU};
}
What is the output of the program? ____________

Analysis

1.      Time spent to solve the problem _____________

2.      Mental workload rate______________
(Rate from 1 to 10
    Where, 1 - very less mental workload used
    10 - very high mental workload used)

3.      General Task Difficulty  _______________
(Rate from 1 to 10
    Where, 1 âĂŞ very easy task
    10 âĂŞ Impossible to solve with only mind)
```

```
11. Solve the below java program and fill the output
    and analysis section.

public class Main {
```

```
    public static void main(String[] args) {

        String s = "";
        boolean b1 = true;
        boolean b2 = false;

        if(b2 = false  |  (21%5)  > 2 )
            s += "x";

        if(b1 || (b2 == true))
            s += "y";

        if(b2 == true)
            s += "z";

        System.out.println(s);
    }
}
```

What is the output of the program? ____________

Analysis

1. Time spent to solve the problem _____________

2. Mental workload rate______________

(Rate from 1 to 10

Where, 1 – very less mental workload used

10 – very high mental workload used)

3. General Task Difficulty _______________

(Rate from 1 to 10

```
Where, 1 âĂŞ very easy task
10 âĂŞ Impossible to solve with only mind)
```

```
12. Solve the below java program and fill the output
    and analysis section.


public class Main {
    public static void main(String[] args) {

        int mask =0;
        int count =0;

        if( ((5<7) || (++count < 10)) | mask++ <10)
            mask = mask +1 ;

        if( (6>8) ^ false)
            mask = mask + 10;

        if( !(mask > 1) && ++count >1)
            mask = mask +100;

        System.out.println (mask);

    }
 }

What is the output of the program? _____________

Analysis
```

```
1.        Time spent to solve the problem ______________

2.        Mental workload rate______________
(Rate from 1 to 10
    Where, 1 - very less mental workload used
    10 - very high mental workload used)

3.        General Task Difficulty ________________
(Rate from 1 to 10
    Where, 1 âĂŞ very easy task
    10 âĂŞ Impossible to solve with only mind)
```

```
13. Solve the below java program and fill the output
    and analysis section.



public class Main {
    public static void main(String[] args) {

        int x =2;
        int y =3;

        if ( (y == x++) |  (x < ++y) ) {
            System.out.println (x + "," + y);
        }

    }
}
What is the output of the program? _____________
```

```
Analysis

1.      Time spent to solve the problem ______________

2.      Mental workload rate_______________
(Rate from 1 to 10
    Where, 1 - very less mental workload used
    10 - very high mental workload used)

3.      General Task Difficulty ________________
(Rate from 1 to 10
    Where, 1 âĂŞ very easy task
    10 âĂŞ Impossible to solve with only mind)
```

```
14. Solve the below java program and fill the output
    and analysis section.

public class Main {

    public static void main(String[] args) {
        int x = 3;
        int y = 5;
        int z = 8;

        x++;
        y++;
        z++;

        String result = (x > 4 && y < 6)
```

```
            ? "Dog" : (x > 6 && z > 8)? "Cat" : "Horse";

        System.out.println(result);
    }
}
```
What is the output of the program? ____________

Analysis

1. Time spent to solve the problem _____________

2. Mental workload rate______________
(Rate from 1 to 10
    Where, 1 – very less mental workload used
    10 – very high mental workload used)

3. General Task Difficulty _______________
(Rate from 1 to 10
    Where, 1 âĂŞ very easy task
    10 âĂŞ Impossible to solve with only mind)

---

15. Solve the below java program and fill the output
    and analysis section.

```
public class Main {
    public static void main(String[] args) {

        int x =42;
        int y =44;
```

```
        System.out.print(7 + 2 + " ");
        System.out.print(5 + foo() + x + 5 + " ");
        System.out.println( x + y + foo());

    }
    static String foo(){

        return "foo";
    }
}
```

What is the output of the program? ____________

Analysis

1. Time spent to solve the problem _____________

2. Mental workload rate______________

(Rate from 1 to 10

Where, 1 – very less mental workload used

10 – very high mental workload used)

3. General Task Difficulty _______________

(Rate from 1 to 10

Where, 1 âĂŞ very easy task

10 âĂŞ Impossible to solve with only mind)

---

16. Solve the below java program and fill the output and analysis section.

(Note: Rectangle object is defined as drawRectangle(x–axis, y–axis, width, height))

```
class MyCanvas extends JComponent {

  public void paint(Graphics g) {
    g.drawRect (10, 10, 10, 10);
    g.drawRect(30, 50, 10, 10);

  }
}
```

Will the two rectangles overlap? ____________

Analysis

1. Time spent to solve the problem _____________

2. Mental workload rate______________
(Rate from 1 to 10
Where, 1 – very less mental workload used
10 – very high mental workload used)

3. General Task Difficulty _______________
(Rate from 1 to 10
Where, 1 âĂŞ very easy task
10 âĂŞ Impossible to solve with only mind)

17. Solve the below java program and fill the output and analysis section.
(Note: Rectangle object is defined as drawRectangle(x–axis, y–axis, width, height))

```
class MyCanvas extends JComponent {
  public void paint(Graphics g) {

      int x1 = 10;
      int y1 = 10;
      int w1 = 130;
      int h1 = 90;

      int x2 = 130;
      int y2 = 50;
      int w2 = 100;
      int h2 = 100;

    g.drawRect (x1, y1, w1, h1);
    g.drawRect(x2, y2, w2, h2);

  }
}
Will the two rectangles overlap? _____________

Analysis

1.        Time spent to solve the problem ______________

2.        Mental workload rate______________
(Rate from 1 to 10
    Where, 1 - very less mental workload used
    10 - very high mental workload used)
```

```
3.      General Task Difficulty ________________
(Rate from 1 to 10
    Where, 1 âĂŞ very easy task
    10 âĂŞ Impossible to solve with only mind)
```

```
18. Solve the below java program and fill the output
        and analysis section.
(Note: Rectangle object is defined as
        drawRectangle(x-axis, y-axis, width, height))



class MyCanvas extends JComponent {

  public void paint(Graphics g) {

      int x1 = 3;
      int y1 = 8;
      int w1 = 20;
      int h1 = 30;

      int x2 = 20;
      int y2 = 30;
      int w2 = 20;
      int h2 = 30;

      x1 = x1 * 2;
      y1 = y1 * 2;

    g.drawRect (x1, y1, w1, h1);
    g.drawRect(x2, y2, w2, h2);
```

```
  }
}
Will the two rectangles overlap? ____________

Analysis

1.      Time spent to solve the problem ______________

2.      Mental workload rate______________
(Rate from 1 to 10
    Where, 1 – very less mental workload used
    10 – very high mental workload used)

3.      General Task Difficulty ________________
(Rate from 1 to 10
    Where, 1 âĂŞ very easy task
    10 âĂŞ Impossible to solve with only mind)
```

```
19. Solve the below java program and fill the output
        and analysis section.
(Note: Rectangle object is defined as
        drawRectangle(x–axis, y–axis, width, height))

class MyCanvas extends JComponent {
  public void paint(Graphics g) {

      int x1 = 20;
      int y1 = 10;
      int w1 = 15;
```

```
        int h1 = 15;

        int x2 = 50;
        int y2 = 50;
        int w2 = 100;
        int h2 = 100;

        x1 = x1/2;
        y1 = y1/2;

        x2 = x2/2;
        y2 = y2/2;

      g.drawRect (x1, y1, w1, h1);
      g.drawRect(x2, y2, w2, h2);
  }
}
Will the two rectangles overlap? ____________

Analysis

1.        Time spent to solve the problem _____________

2.        Mental workload rate______________
(Rate from 1 to 10
    Where, 1 – very less mental workload used
    10 – very high mental workload used)

3.        General Task Difficulty _______________
(Rate from 1 to 10
    Where, 1 âĂŞ very easy task
```

```
10 âĂŞ Impossible to solve with only mind)
```

```
20. Solve the below java program and fill the output
      and analysis section.
(Note: Rectangle object is defined as
      drawRectangle(x–axis, y–axis, width, height))


class MyCanvas extends JComponent {

  public void paint(Graphics g) {

      int x1 = 10;
      int y1 = 10;
      int w1 = 15;
      int h1 = 15;

      int x2 = 15;
      int y2 = 20;
      int w2 = 15;
      int h2 = 10;

      int x3 = 10;
      int y3 = 45;
      int w3 = 25;
      int h3 = 10;

    g.drawRect (x1, y1, w1, h1);// Rectangle1
    g.drawRect(x2, y2, w2, h2);// Rectangle2
    g.drawRect(x3, y3, w3, h3);// Rectangle3
```

```
  }
}
```

Which two rectangles overlap? ____________

Analysis

1. Time spent to solve the problem _____________

2. Mental workload rate______________

(Rate from 1 to 10

Where, 1 – very less mental workload used

10 – very high mental workload used)

3. General Task Difficulty _______________

(Rate from 1 to 10

Where, 1 âĂŞ very easy task

10 âĂŞ Impossible to solve with only mind)

Background Questionnaire

Background Questionnaire - September 2016

1. **Name**: ________________________________

2. **Age:** ________________________

3. **Email:** ______________________________________

4. **Gender**: ☐ Male ☐ Female

5. **Years of higher education (university and up)**___________________

6. **I have degrees in (including degree currently pursuing):**

   ☐ Computer Science

   ☐ Software Engineering

   ☐ Other computing-related field (specify): __________________

   ☐ Other engineering field (specify): __________________

   ☐ Other field (specify): __________________

7. **Highest degree obtained:** ____________________________

8. **Highest degree obtained year:** __________________

9. **Years/months of experience in Java Programming:** ____________

Note: Age, gender and level of education will be anonymously linked with the EEG data. Other personal information will be kept secure initially and will be destroyed after 3 months.

Exit Questionnaire

Exit Questionnaire - September 2016

Name: ____________________________

(information here will be anonymized prior to linking with EEG data)

Observations about the experimental tasks

1. Difficulty: Overall I found the tasks to be

☐ Very difficult ☐ Difficult ☐ Somewhat difficult ☐ Somewhat easy ☐ Easy

☐ Very easy

2. Adequacy: I found the explanations and documentation provided to be

☐ Comprehensive ☐ Adequate ☐ Inadequate

3. Enjoyment: I found the tasks to be

☐ Very enjoyable ☐ Enjoyable ☐ Somewhat enjoyable ☐ Somewhat boring

☐ Boring ☐ Very boring

4. Rank each of these problems from easy to difficult by assigning them numbers from 1 to 10. If you think some of the problems are equally difficult, you can assign them the same number.

Question	Rank from 1 (Easiest) to 10 (Hardest)
#1	
#2	
#3	
#4	
#5	
#6	
#7	
#8	
#9	

5. Comments

NASA TLX Questionnaire

NASA TLX - RATING SHEET
September – 2016

Subject ID: ________________________________ Task ID: ____________________________________

Put an "X" on each of the six scales at the point which matches your experience.

PERFORMANCE

Good Poor

MENTAL DEMAND

Low High

PHYSICAL DEMAND

Low High

TEMPORAL DEMAND

Low High

EFFORT

Low High

FRUSTRATION

Low High

SOURCES-OF-WORKLOAD COMPARISON BOXES

In order to assess the relative importance of the factors in the rating scale, we ask you to choose the most important factor in a pair of rating scale titles.

In each box circle the scale title that represents the more important contributor to workload for the specific task(s) you performed in this experiment.

Effort or Performance	Temporal Demand or Frustration	Temporal Demand or Effort
Physical Demand or Frustration	Performance or Frustration	Physical Demand or Temporal Demand
Physical Demand or Performance	Temporal Demand or Mental Demand	Frustration or Effort
Performance or Mental Demand	Performance or Temporal Demand	Mental Demand or Effort
Mental Demand or Physical Demand	Effort or Physical Demand	Frustration or Mental Demand

RATING SCALE INSTRUCTIONS

In order to assess the experience you had during the tasks you carried out, we are asking you to fill out a sheet of rating scales. In the most general sense, we are examining the "workload" you experienced. Each line on the rating scale has two end point descriptors that describe the scale. Please put an "X" on each of the six scales at the point which matches your experience. The scales we used are described below. Note that "Performance" scale goes from "good" on the left to "bad" on the right while the other scales go from "low" in the left to "high" on the right. Please consider your responses carefully distinguishing among the different task conditions. Consider each scale individually. Your ratings will play an important role in evaluating our study. Your participation is essential to the success of this experiment and is greatly appreciated by us.

RATING SCALE DEFINITIONS

Title	**Endpoints**	**Descriptions**
PERFORMANCE	*Good/Poor*	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfies were you with your performance in accomplishing these goals?
MENTAL DEMAND	*Low/High*	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	*Low/High*	How much physical activity was required (e.g., pushing, pulling, controlling, activating, holding a position etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	*Low/High*	How much time pressure did you feel due to the rate or pace at which the task or tasks elements occurred? Was the pace slow and leisurely or rapid and frantic?
EFFORT	*Low/High*	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVEL	*Low/High*	How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?

NASA TLX - Calculation

APPENDIX A: Sample Application of the NASA-TLX.

EXAMPLE:

RESPONSES. THE PRIMARY DIFFICULTY MANIPULATION IS THE INTER-STIMULUS INTERVAL (IS!) - (TASK 1 = 500 msec. TASK 2 = 300 msec)

SIGNIFICANT SOURCE OF WORKLOAD VARIATION IN THESE TASKS

PD / MD	TD / PD	TD / FR
TD / MD	OP / PD	TD / EF
OP / MD	FR / PD	OP / FR
FR / MD	EF / PD	OP / EF
EF / MD	TD / OP	EF / FR

TALLY OF IMPORTANCE SELECTIONS

MD III	=	3
PD	=	0
TD IIIII	=	5
OP I	=	1
FR III	=	3
EF III	=	3
SUM	=	15

RATING SCALES:

TUDE OF EACH FACTOR IN THE TASK YOU JUST PERFORMED

DEMANDS	RATINGS FOR TASK 1:			RATING		WEIGHT		PRODUCT
MD	LOW	I _x_______I	HIGH	30	X	3	=	90
PD	LOW	I x________I	HIGH	15	X	0	=	0
TD	LOW	I _______x__I	HIGH	60	X	5	=	150
OP	EXCL	I ___x______I	POOR	40	X	1	=	40
FR	LOW	I _x_______I	HIGH	30	X	3	=	90
EF	LOW	I ____x_____I	HIGH	40	X	3	=	120
				SUM			=	490
				WEIGHTS (TOTAL)			=	15
				MEAN WWL SCORE			=	$\boxed{32}$

DEMANDS	RATINGS FOR TASK 2:			RATING		WEIGHT		PRODUCT
MD	LOW	I _x_______I	HIGH	30	X	3	=	90
PD	LOW	I _x_______I	HIGH	25	X	0	=	0
TD	LOW	I _______x__I	HIGH	70	X	5	=	350
OP	EXCL	I _____x____I	POOR	50	X	1	=	50
FR	LOW	I ____x_____I	HIGH	50	X	3	=	150
EF	LOW	I _x_______I	HIGH	30	X	3	=	90
				SUM			=	730
				WEIGHTS (TOTAL)			=	15
				MEAN WWL SCORE			=	$\boxed{49}$

Consent form used in experiment

UNIVERSITY OF KENT

FORM OF CONSENT TO TAKE PART IN A RESEARCH PROJECT

CONFIDENTIAL

Title of project / investigation: Cognitive Task Difficulty Analysis using EEG and Data Mining

Brief outline of project, including an outline of the procedures to be used:

Outline of the project

The project is intended to measure and classify task difficulty of program comprehension tasks using EEG and Data Mining. During the EEG recording session, the subjects will be presented Java programs of various difficulties and solve the programs. The different areas of activation and different frequency band originated in brain will be collected and then analysed. Machine learning algorithm will be applied on the collected data for best feature selection to predict task difficulty.

Preparation of the subject

Each session will start as follows:

The Emotive EEG device will be placed on the subject's head. The device is placed on the head and communicates with the computer system wirelessly. Saline water will be used to improve the signal conductivity. This procedure is harmless and painless. More details about the device can be found in the following link: http://emotiv.com/epoc/

EEG is a widely used medical technique for recording brain activity. There is no electrical current transmission to the subject and only currents produced by the brain are recorded.

An outline of the experimental procedure

Subjects will be asked to solve Java program displayed on the screen. The brain signal generated will be capture by the EEG device and filtered and analysed using Fourier Transform technique. Each session will last up to an hour including the experimental setup. The subject will remain seated during the whole experiment.

The subjects will be able to take rests if they need to.

The identity of all volunteers will be kept anonymous and private at all times. Volunteers can withdraw at any time (even during experiments) without having to give any reason.

I, ... *(**participant's** full name) agree to take part in the above named project / investigation, the details of which have been fully explained to me and described in writing. I confirm that I am not taking medication that may affect alertness and have no history of stroke, epilepsy, or neural disorders.

Signed .... Date ....

(Participant)

I, ... *(**Investigator's** full name) certify that the details of this project / investigation have been fully explained and described in writing to the subject named above and have been understood by him / her.

Signed .... Date ....

(Investigator)

*Please type or print in block capitals

Volunteer Information

UNIVERSITY OF KENT

SCHOOL OF COMPUTING

VOLUNTEERS NEEDED

FOR THE EXPERIMENT

"Cognitive Task Difficulty Analysis using EEG and Data Mining"

Experimental task

You will be asked to solve different Java programming of various difficulties and fill out NASA TLX survey based on your insight on the task difficulty. The solution to the tasks will be input via the keyboard. Prior to the first task and in-between each task you will be given time to relax yourself.

Brain activity will be recorded during the experiment. Recording session will be approximately up to an hour and you will be paid £15 for your time upon completion of the data collection.

Experiments will take place in the Medway campus or in SW08 of Canterbury campus.

The procedure is painless and harmless. You will have right to stop the experiment at any time. The identity and privacy of all volunteers will be protected, and recorded data will be anonymous.

Experiments will start in early July.

General requirements

- Age 18-50
- Participants who can sit ALERT during the length of the experiment
- Participants who are not taking medication that may affect alertness
- Participants with no history of stroke, epilepsy or panic attack
- Participants who have sufficient Java programming experience.

Contact person(s)

If you would like to know more about the experiment, contact Mrs Aruna Duraisingam (ad543@kent.ac.uk) or Dr Palaniappan Ramaswamy (r.palani@kent.ac.uk).

APPENDIX J

Classification - Matlab Code

```
clc
clear all
close all



% Author: Aruna Duraisingam
% Date : 20-Aug-2016
% Main script to import the inputand target data from
 Excel sheet and
% perform 10 fold classification based on task difficulty
 level (Easy,
% medium and difficult).



% import input and target data from excel sheet

```

```
data = importdata('ClassifyData/classification_easydiff.xlsx');

% define the input and target data. The size of the
 inputdata and the
% target data should be same. The size should be same
 between class records
% as well.

inp =data.InputData;
target = data.TargetData;



% Cross validation function automatically splits into 10
 folds data
% (default) and perfro recursively based on the input and
 the target data.
% fun is a function where the feedforward neural network
 is designed and
% evaluated.

vals = crossval(@(XTRAIN, YTRAIN, XTEST,
 YTEST)networkFunction(XTRAIN, YTRAIN, XTEST, YTEST),
  inp, target);
```

```

% Author: Aruna Duraisingam
% Date : 20-Aug-2016
% Function to execute each fold train and test data
 (recursively) and
% classify task difficulty level and compute
 classification success rate.
```

```
function testval = networkFunction(XTRAIN, YTRAIN, XTEST,
 YTEST)

% creates a two layer network with 10 hidden layers.

% Default training function trainlm.

net = feedforwardnet(10);

% Initializes the weight and bias
net = init(net);

% modified the transfer function of the layers in netwrok
net.layers{2}.transferFcn = 'logsig';
net.layers{1}.transferFcn = 'logsig';

net.divideParam.trainRatio = 75/100;
net.divideParam.valRatio = 15/100;
net.divideParam.testRatio = 15/100;

net.trainParam.epochs = 100;

% Train the network. Defualt performance function 'mse'
 (mean Squared
% Error)
[net, tr] = train(net, XTRAIN', YTRAIN');

% Network responses after trained and validated
% simulate a network over  input range
```

```
output = net(XTEST');
% find which have highest probability value
[ maxout ,classOut] = max(output',[],2);

% convert  into a format that can be compared with maxindex
[ testout ,testIndex] = find(YTEST);

% Correctly classified samples divided by the classified
 samples.
% Inconclusive results are not counted
% Check the success of the classifier
cp = classperf(testIndex, classOut);

%
%       figure, plotperform(tr)
%       figure, plottrainstate(tr)



% get the correctrate
testval = (cp.CorrectRate) * 100;
end
```