9-19-2019

Cognitive Load Detection For Advanced Driver Assistance Systems

Prathamesh Ayare
University of Windsor

COGNITIVE LOAD DETECTION FOR ADVANCED DRIVER ASSISTANCE SYSTEMS.

by

Prathamesh Ayare

A Thesis
Submitted to the Faculty of Graduate Studies
through the Department of Electrical and Computer Engineering
in Partial Fulfilment of the Requirements for
the Degree of Master of Applied Science at the
University of Windsor

Windsor, Ontario, Canada

Cognitive Load Detection for Advanced Driver Assistance Systems

by

Prathamesh Ayare

APPROVED BY:

J. Ahamed
Department of Mechanical, Automotive and Materials Engineering

M. Abdelkhalek
Department of Electrical and Computer Engineering

B. Balasingam, Co-Advisor
Department of Electrical and Computer Engineering

F. Biondi, Co-Advisor
Department of Kinesiology

September 19, 2019

Declaration of Co-Authorship/ Previous Publication

Co-Authorship

I hereby declare that this thesis incorporates material that is result of joint research, as follows: Chapters 2 and 3 of this thesis were co-authored with professor Balasingam professor Milne, and professor Biondi who provided supervision and guidance during the research and writing process. In all cases, the key ideas, primary contributions, experimental designs, data analysis, interpretation, and writing were performed by the author. I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my thesis, and have obtained written permission from each of the co-author(s) to include the above material(s) in my thesis.

Previous Publication

This thesis includes two original papers that have been previously published/submitted for publication in peer reviewed journals, as follows:

Thesis chapter	Publication title/full citation	Publication status
2	P. Ayare, B. Balasingam, F. Biondi and k.Milne, “On the Feasibility of Cognitive Load Detection through Pupil Dilation Measurements in ADAS”, submitted, IEEE Transactions in Human Machine Systems, August 2019.	Submitted
3	P. Ayare, B. Balasingam, F. Biondi and k.Milne, “ Comparison of Cognitive Load Classification Based On Pupil Dilation and DRT Reaction Time”, submitted, IEEE Transactions in Human Machine Systems, August 2019.	Submitted

I certify that I have obtained a written permission from the copyright owner(s) to include the above published material(s) in my thesis. I certify that the above material describes work completed during my registration as a graduate student at the University of Windsor.

General

I declare that, to the best of my knowledge, my thesis does not infringe upon anyones copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

Abstract

In this thesis, we investigate cognitive load detection and classification based on minimally invasive methods. Cognitive load detection is crucial for many emerging applications such as advanced driver assistance systems (ADAS) and industrial automation. Numerous studies in the past have reported several psychological measures, such as eye-tracking, electrocardiogram (ECG), electroencephalogram (EEG), as indicators of cognitive load. However, existing physiological features are invasive in nature. Consequently, the objective of this study is to determine the feasibility of non-invasive features such as pupil dilation measurements low-cost eye-tracker with minimal constraints on the subject for cognitive load detection. In this study, data from 33 participants were collected while they underwent tasks that are designed to permeate three different cognitive difficulty levels with and without cognitive maskers and the following measurements were recorded: eye-tracking measures (pupil dilation, eye-gaze, and eye-blinks), and the response time from the detection response task (DRT). We also demonstrate the classification of cognitive load experienced by humans under different task conditions with the help of pupil dilation and reaction time. Developing a model that can accurately classify cognitive load can be used in various sectors such as semi-autonomous vehicles and aviation. we have used a data fusion approach by combining pupil dilation and DRT reaction time to determine if the classification accuracy increases. Further, we have compared the classifier with the highest classification accuracy using data fusion against the accuracy of the same classifier with only one feature (pupil dilation; reaction time) at a time.

Acknowledgements

First of all, I would like to express my deep sense of gratitude to my supervisor Dr. B. Balasingam, who offered me an opportunity to be a graduate student and conduct research under his guidance. Further, I would like to thank Dr. Francesco Biondi for his advice and suggestions throughout the course of this thesis. They gave me all the support and confidence needed to complete this thesis. I thank Dr. Balasingam for spending countless hours discussing my work, the papers and this thesis.

I would like to thank all the staff members of our department for their kind support. Especially, I thank Andria Ballo and Connie laquinta for their constant help and support.

I thank the fellow graduate students in the Human Machine Interaction laboratory (HMS Lab), who gave me a perfect research environment to work in with all the fun and offered support whenever needed. I would also like to thank all the graduate and undergraduate students who participated in my research.

I take this opportunity to thank Dr. Kevin Milne for his support. I thank my parents for their endless love and support. And, finally, I would like to thank my friends, for being a part of my success.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In this manuscript-style thesis, we explore the research and development of technology in the realm of cognitive load detection using psychological measurements presented by the author as a collection of their previously submitted works. Cognitive load detection and classification using psychological measures itself is a broad topic which boasts an increasing number of practical applications in Semi-autonomous vehicles, aviation, and other industries. One specific application of cognitive load detection and classification is in semi-autonomous vehicles to keep the driver safe during unusual circumstances.

Autonomous and semi-autonomous industries continuously strive to make the driving experience as safe as possible. Compared to full manual control, vehicles now come with advanced features such as adaptive cruise control, automatic lane change, the anti-braking system which assist the driver to have a safe and smooth ride. These features are termed as advanced driver assisted systems (ADAS) [8]. ADAS is formally defined in the later section. Even though these features help ensure safety, the choice of activating these features remains with the driver. The driver being a human is usually unaware of his or her cognitive state. High cognitive load causes mental fatigue which deteriorates the ability of the driver to drive with precaution. Such a cognitive state can result in a condition that might be dangerous for both the fellow passengers on the road as well as the driver. Failure to activate the ADAS under such circumstances can result in tragic incidents. ADAS systems can be designed to adapt and activate themselves with a driver's cognitive state making the driving experience even more safer [11]. To make reliable and accurate, cognitive load detection and classification models based on psychological measurements i.e. reaction time [9] and pupil dilation [7] which can be applied in the field of semi-autonomous vehicles and other industries will be explored later in the chapters. As such the author has dedicated his research and time to this end and has striven to investigate minimally invasive techniques to help detect and classify cognitive load depending on the difficulty of the task.

1.1 Organization of the Thesis

The author has elected to present this thesis structured according to the manuscript format rather than the traditional format. That is, the chapters to follow consist of manuscripts previously written and submitted by the author, with the first authorship, are included in this thesis as written at the times of their submissions, in chronological order, with alterations to format and slight modifications to content to maintain a cohesive thesis structure. As prescribed by the manuscript format, abstracts have also been omitted. The author believes that by virtue of the chosen format their thought process, understanding of the research topic and its place in the modern world, and journey toward producing increasingly meaningful contributions, are far more accurately conveyed as a story told through a collection of chronological works.

While a traditional thesis commonly contains a general literature review and problem statement, the author has chosen to forego these sections in the traditional sense. The reader will find that each of the manuscript chapters provides their own introductions which serve the purpose of familiarizing the reader with both the context of the research and relevant literature - however, the content discussed in each reflects the knowledge and understanding of the topic at their respective times of writing. As such, the next section will present the key points from each manuscript chapter introduction as well as some additional thoughts and findings at the time of writing. Each of the manuscript chapters also contain a section describing the problem to be addressed by the research. To include a general literature review and problem statement in this thesis would be to introduce unnecessary redundancy. It must be noted, however, that some amount of redundancy will persist throughout the manuscript chapters as a consequence of each being originally written as their own, standalone entities.

The remainder of this thesis is organized as follows: Chapters 2 provides brief description of the data collection procedure, type of data collected, experimental design, tasks, descriptive and inferential statistical analysis of the collected data, comparison of dual-task and single-task performance on cognitive load and feasibility of pupil dilation as a measure of cognitive load. A conceptual understanding of each is necessary to understand the subject matter of Chapter 3. In Chapter 3, where the first author demonstrates the classification of cognitive load using three different classifiers i.e. support vector machine, logistic regression, and k-means with data fusion. In the later section, the first author compares the above-mentioned classifiers based on classification accuracy. Chapter 3 also describes the use of a new method for cognitive load detection based on pupil dilation and reaction time. Finally, the author concludes the findings of this research along with future work.

1.2 Understanding Advanced driver assistance systems

The world health organization reported a global status on road safety in 2018 claiming 1.35 million deaths from car accidents [3]. A 2016 global report from safer America showed that 95% of all vehicles involved in fatal car accidents were passenger cars or light trucks and 50% of fatalities in car accidents were drivers [1]. The National Highway Traffic Safety Association from different countries regulate and enforce safety standards placed on automobile manufacturers. Due to this manufacturers are competing to offer a cutting edge system that assists drivers in safe and accident-free driving. Such systems are known as advanced driver assistance systems (ADAS). The ADAS is an intelligent safety system that tries to improve road safety in terms of crash avoidance, crash severity mitigation and protection, and automatic post-crash notification of collision [2]. There are various driver assistance systems each working to provide a different feature. Some are critical to safety whereas some help driver avoids minor accidents. Examples of ADAS are automatic cruise control, automatic lane change, lane keep assist and collision braking mitigation system [10].

1.3 Pupil dilation and Its applications

Pupil dilation is usually measured using eye trackers. Eye tracker uses infrared cameras to capture pupil dilation. With the recent development in technology, pupil dilation can be measured using low-cost portable cameras thus making pupil dilation useful in a range of applications. Researchers have suggested that pupil dilation contains a lot of information regarding human behavior, health [6] and emotions [12]. Human behavior, health, and emotions can be predicted from the pupil dilation which has found a variety of applications like:

1. Interpretation and analysis of fMRI data: Siegle, stenger, konecky and carter [13] compared the time course of pupil dilation with that of fMRI signal in the middle frontal gyrus during a digit sorting task to suggest that activity in that area indexed the working memory subtask of digit sorting.

2. Lie detection: Wang, spezio, and camerer [15] demonstrated that pupil dilation is proportional to the size of the lie, which can be due to two reasons i.e. lying

involves more complicated process or simply due to guilt.

3. cognitive workload detection: Researchers examined the effect of increasing cognitive workload tasks on pupil dilation and found out that pupil dilation increases with task difficulty and hence can be a good measure of cognitive load [10].

1.4 Detection Response Task overview

Detection response task (DRT) is a method that evaluates the attentional aspect of cognitive load in drivers. Response time and the missed rate is measured using DRT. Response time is interpreted as the attentional effect of cognitive load [5]. DRT consists of a different form of stimulus i.e. visual, auditory and tactile and the response is in the form of micro-switch. In visual DRT the driver is presented with a visual target in the form of a red dot. The driver has to press the microswitch every time the visual target is presented. In auditory DRT the driver is presented with an audio target in the form of the auditory signal. The driver has to press the microswitch every time the hear auditory signal. In tactile DRT the driver is presented with a stimulus in the form of vibrations produced by a small vibration generator device. The driver has to press the microswitch every time they feel a vibration produced by the device.

All the types of DRT measure response time and missed rate in a similar fashion. Response time is measured as time from the stimulus onset until the response onset [5]. Missed rate is measured as no response is given within 100 - 2500 milliseconds after stimulus onset [5]. Many researchers have used DRT to measure the cognitive load experienced by the driver [4]. Longer reaction time indicate a higher cognitive load [14]. DRT is an accepted standard to measure cognitive load in the driving environment [9].

1.5 Bibliography

[1] "Road traffic deaths," Mar 2016. [Online]. Available: https://www.who.int/gho/road_safety/mortality/en/ pages 4

[2] "European commission - traffic safety synthesis 2016," Feb 2017. [Online]. Available: https://www.nrso.ntua.gr/european-commission-traffic-safety-synthesis-2016/ pages 4, 8

[3] "100 car accident statistics for 2019," Oct 2018. [Online]. Available: https://safer-america.com/car-accident-statistics/#Global pages 3

[4] M.-P. Bruyas and L. Dumont, "Sensitivity of detection response task (drt) to the driving demand and task difficulty," 2013. pages 5

[5] I. O. for Standardization, "Road vehiclestransport information and control systemsdetection–response task (drt) for assessing attentional effects of cognitive load in driving." 2016. pages 5, 10, 14, 25, 49, 52

[6] J. H. Growdon, K. Graefe, M. Tennis, D. Hayden, D. Schoenfeld, and S. H. Wray, "Pupil dilation to tropicamide is not specific for alzheimer disease," *Archives of neurology*, vol. 54, no. 7, pp. 841–844, 1997. pages 4

[7] J. Hyönä, J. Tommola, and A.-M. Alaja, "Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks," *The Quarterly Journal of Experimental Psychology*, vol. 48, no. 3, pp. 598–612, 1995. pages 2, 49

[8] M. Lu, K. Wevers, and R. Van Der Heijden, "Technical feasibility of advanced driver assistance systems (adas) for road traffic safety," *Transportation Planning and Technology*, vol. 28, no. 3, pp. 167–187, 2005. pages 1

[9] T. A. Ranney, G. Baldwin, L. A. Smith, E. N. Mazzae, and R. S. Pierce, "Detection response task (drt) evaluation for driver distraction measurement application," Tech. Rep., 2014. pages 2, 5, 49

[10] A. Shaout, D. Colella, and S. Awad, "Advanced driver assistance systems-past, present and future," in *2011 Seventh International Computer Engineering Conference (ICENCO'2011).* IEEE, 2011, pp. 72–82. pages 4, 5, 8

[11] W. Sheng, Y. Ou, D. Tran, E. Tadesse, M. Liu, and G. Yan, "An integrated manual and autonomous driving framework based on driver drowsiness detection," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems.* IEEE, 2013, pp. 4376–4381. pages 2

[12] G. J. Siegle, S. R. Steinhauer, C. S. Carter, W. Ramel, and M. E. Thase, "Do the seconds turn into hours? relationships between sustained pupil dilation in response to emotional information and self-reported rumination," *Cognitive Therapy and Research*, vol. 27, no. 3, pp. 365–382, 2003. pages 4

[13] G. J. Siegle, S. R. Steinhauer, V. A. Stenger, R. Konecky, and C. S. Carter, "Use of concurrent pupil dilation assessment to inform interpretation and analysis of fmri data," *Neuroimage*, vol. 20, no. 1, pp. 114–124, 2003. pages 4

[14] D. L. Strayer, J. M. Cooper, J. Turrill, J. Coleman, N. Medeiros-Ward, and F. Biondi, "Measuring cognitive distraction in the automobile," 2013. pages 5

[15] J. T.-y. Wang, M. Spezio, and C. F. Camerer, "Pinocchio's pupil: using eyetracking and pupil dilation to understand truth telling and deception in sender-receiver games," *American Economic Review*, vol. 100, no. 3, pp. 984–1007, 2010. pages 4

Chapter 2

On the Feasibility of Cognitive Load Detection through Pupil Dilation Measurements in ADAS

2.1 Introduction

Semi-autonomous vehicles are rapidly taking leverage over manually controlled vehicles. The important factors contributing to this changeover are the advanced driver assistance systems (ADAS) [25] which aims to facilitate driving with minimal effort by the human driver. The ADAS based intelligent safety systems could improve road safety in terms of crash avoidance, crash severity mitigation and protection, and automatic post-crash notification of collision. Further, ADAS could be useful as an integrated in-vehicle or infrastructure-based systems which contribute to all of these crash phases [6]. ADAS takes input from the driver and various in-vehicle systems to produce scheduled outputs such as automatic braking, collision warning and lane change assist which are critical to the safety of drivers [25]. The choice of activating ADAS remains with the driver. It is estimated that although ADAS successfully reduced road accidents by 25% [16], studies have shown that one of the important reasons for ADAS failure is due to human error. Crash data studies have found that driver error and other human factors contribute to as much as 93% of vehicle crashes [9]. Human error can occur due to lack of training on how to use ADAS, the complexity of ADAS, change of human behavior and sole attention on ADAS rather than primary task i.e. driving [10]. Instead of providing full control to the driver, ADAS can be designed to adapt and trigger based on feedback received from the driver's cognitive state. Consequently, cognitive load detection has become one of the active research topics of the recent past. Tianyi Hong and Huabiao Qin [15] proposed to detect drowsiness through the percentage of eye closing data, i.e., the number of blinks, through steering mounted camera and depending on the level of drowsiness a warning message was displayed or the vehicle was slowed down and stopped thus avoiding a crash or any unusual circumstances. Humans have limited capacity and the ability to focus attention deteriorates under conditions of high load on cognitive control processes such as working memory [17]. Cognitive load delays a driver's response to critical events [9]. Under conditions of high cognitive load, failure of ADAS due to human error can result in undesired incidents such as accidents.

One of the crucial steps in developing a cognitive state-adaptive ADAS is measuring the cognitive load experienced by the driver [9, 11]. Unlike physical load, the cognitive load experienced by an individual cannot be determined through direct measurements. Researchers have suggested that cognitive load can be detected using analytical and empirical methods [27]. Analytical methods are aimed at evaluating the cognitive load by collecting analytical data with methods such as mathematical models and task analysis. Empirical methods involve estimating the cognitive load by collecting subjective data using rating scales, performance data using primary and secondary task techniques and psychological data using psychological techniques. Psychological techniques are based on the assumption that changes in cognitive functioning are indicated by psychological variables [23]. Heart rate, heart rate variability, reaction time and pupil dilation are some examples of psychological variables. Detection response task (DRT) is one of the accepted standards for measuring reaction time among drivers capturing the attentional aspect of cognitive load [13]. Literature suggests that reaction times are longer with higher cognitive load imposed by the task [28]. Although DRT is an accepted standard for measuring cognitive load, it needs to be in contact with the driver/participant under study; hence, DRT is considered an invasive measure of cognitive load detection. What is primarily preferred in ADAS and other human-machine automation systems are non-invasive measures of which eye-tracking is a good example. The cognitive functioning of the human brain can be detected through eye-tracking measurements such as pupil dilation, eye gaze, and eye blinks. Pupil dilation can dilate up-to 0.5mm in response to cognitive processing stimuli [8]. Researchers have suggested that any sensory movement — whether tactile, auditory or gustatory — tigers pupillary reflex dilation. The mental process, emotional effort, and motor actions also evoke the pupil dilation [8]. Past studies have shown that the mean and variance of the pupil dilation increases with cognitive difficulty [18]. It was also shown that eye-tracking can be used for detecting and tracking transient changes in the pupil dilation for multiple levels of cognitive difficulty [18]. Short-duration studies involving pupil dilation suggest that while information is received into the memory pupil dilates slightly, dilation increases when the information is processed and constricts when information is retrieved. For the long-duration task, the peak pupillary dilation is consistently higher than the short duration task but the constriction during memory retrieval is almost similar in both the conditions [7].

Despite the challenges in using pupil dilation as a measure of cognitive load, the technical advances in optical sensing and artificial intelligence continue to improve the feasibility of this becoming a reality in the future. With the recent development in the field of eye-tracking pupil dilation can be measured through low-cost cameras with high efficiency. Researches have come up with an accurate and real-time estimation of pupil size through portable cameras such as webcams [24] which can be easily installed in vehicles. With successful detection of the cognitive state through non-invasive measures, such as eye tacking, it can be possible to develop more reliable ADAS.

In this chapter, we compare the measured pupil dilation against the following measures of cognitive load:

- *Subjective measures:* The NASA task load index (NASA-TLX) [14], a retrospective set of questionnaires, is one of the well known subjective measures of cognitive load.

- *Standardized measures:* The International Standards Organization (ISO) has standardized DRT as an acceptable measure of cognitive workload (ISO 17488, 2016). The following DRT measures were recorded for comparative analysis:

    1. *Response time:* The time it took for the participant to respond to a stimuli (administered in the form of buzzer) by pressing a button. It is expected that as the cognitive load experienced by the participant increases, so will the response time.

    2. *Missed trials:* The number of times the participant failed to respond to the stimuli. It is expected that as the cognitive load experienced by the participant increases, so will the missed trials.

- *Performance measures:* Here, the accuracy of the $n-$back experiment is considered to be an indicator of cognitive load.

Based on the analysis performed, the following observations were made in this chapter:

1. *Pupil dilation increased with cognitive load:* When tested among four difficulty levels, pupil dilation showed a statistically significant difference in 3 out of 6 pairs of cognitive load.

2. *Reaction time increased with cognitive load:* When tested among four difficulty levels, reaction time showed a statistically significant difference in 2 out of 6 pairs of cognitive load.

3. *Demonstration of multiple measures for cognitive load detection:* Cognitive load was detected by combining the following measures:

   • Subjective measures: NASA-TLX

   • Standardized measures: (a) reaction time and (b) total number of missed trials

   • Psychological measures: Pupil dilation

   • Performance measures: $n-$back accuracy

4. *DRT acted as an additional cognitive load:* This observation is statistically confirmed using pupil dilation. Further, this was confirmed by NASA-TLX as well as $n-$back accuracy (performance measure). We demonstrate the difference between cognitive load experienced by an individual performing a single task vs multitasking through pupil dilation.

The rest of this chapter is organized as follows: Section 3.2 provides the description of the procedure followed for participant recruitment, apparatus used, tasks and procedure followed for conducting the experiment; Section 2.3 consists of data analysis using descriptive statistics such as plotting the mean and standard deviation of collected data; also elaborates inferential statistical techniques used for an in-depth analysis of collected data to validate the findings of descriptive statistics. and Section 3.6 concludes this chapter.

2.2 Data Collection Setup

In this Section, we summarize the data collection setup and procedures.

Figure 2.1: **The Gazepoint (GP3) eye-tracking system** [**3**]. The pupil dilation data were recorded using the GP3 eye-tracking system without imposing any physical constraints on the head; the participants were asked to focus on the '+' sign displayed on the screen.

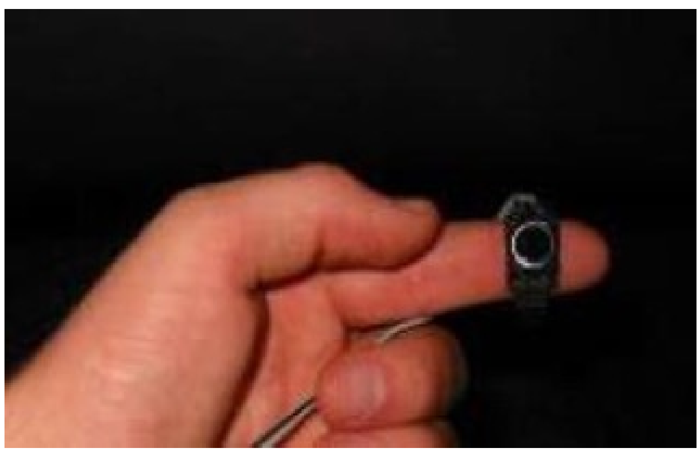
1. **Subjects:** 33 participants ranging in age from 18 to 30 years (M = 22 , SD = 3) were recruited for this study. All the participants were students (18 undergraduate students and 15 graduate students) at the University of Windsor; The solicitations were announced in classrooms and via e-mail circulation at the University of Windsor. Participants received a $20.00 gift card that was announced in the solicitations.

2. **Apparatus:** The following two apparatus were used to collect psychological data during the experiments.

   (a) *Eye-Tracker:* The Gaze-Point (GP3) eye-tracking system [3] was used to collect the following eye-tracking data: pupil dilation, eye-gaze fixations and eye-blinks (Figure 2.1).

(b) *DRT Recorder.* Reaction time was collected through Detection Response Task (DRT) [2]. The DRT device had a stimulus in the form of tactile vibration generator and response through a microswitch (Figure 2.2).

(a) DRT stimulus (vibrations)



(b) DRT response (push-button switch) [13]

Figure 2.2: **DRT stimulus and response.** The DRT stimulus comes in the form a a vibration; in response to each vibration, the participant is required to press the push-button switch shown in (b). The time between the start of the DRT vibration and the response is measured in milliseconds.

3. **Tasks:** The participants had to perform two types of tasks:

   (a) *Detection response task*, referred from hereafter as the *DRT task*. Here, the participant has to press a button in response to the vibration produced

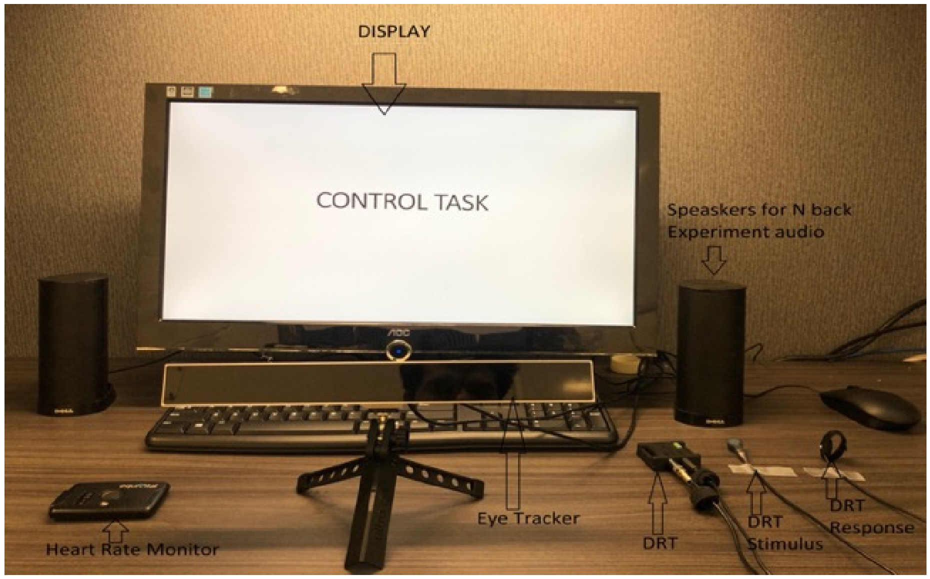


Figure 2.3: **Experiment Setup:** The devices used for data collection.

by the DRT device.

(b) *Delayed digital recall task*, referred from hereafter as the *$n-$back task*. The details of the *$n-$back task* are given next.

The $n-$back task has a serial presentation of a stimulus in the form of audio (series of numbers) spaced approximately one second apart which involves the storage and continual updating of information in working memory [1]. The $n-$back tasks were divided into three stages with increasing difficulty:

- Zero-back: Participants had to repeat out loud same number they just heard (see Table 2.1 for sample response).
- One-back: Participants had to repeat out loud one number previous to the number they just heard (see Table 2.2 for sample response).
- Two-back: Participants had to repeat out loud two numbers previous to the number they just heard (see Table 2.3 for sample response).

The duration of each $n-$back task was approximately three minutes during which approximately seventy $n-$back responses are collected. The first ten stimuli and the expected responses of the $0-$back, $1-$back and $2-$back tasks are listed in Tables 2.1, 2.2 and, 2.3 respectively.

Table 2.1: **Zero-back tasks:** Stimuli and the expected response (first ten out of 72 numbers are shown)

stimulus	$0-$back
8	8
7	7
4	4
5	5
2	2
3	3
1	1
9	9
6	6
0	0

Table 2.2: **One-back tasks:** Stimuli and the expected response (first ten out of 68 numbers are shown)

stimulus	$1-$back
7	-
3	7
6	3
4	6
0	4
5	0
8	5
1	8
9	1
2	9

The $n-$back experiment was divided into two stages:

(a) Dual-Experiment: The participant had to perform the $n-$back task while performing the DRT.

Table 2.3: **Two-back tasks:** Stimuli and the expected response (first ten out of 67 numbers are shown)

stimulus	2−back
5	-
3	-
4	5
8	3
0	4
7	8
1	0
9	7
6	1
8	9

(b) Single-Experiment: Participants had to perform the $n-$back task without the DRT.

4. **Procedure:** Each participant completed two sessions each lasting approximately twenty minutes. Each session was run by a script that causes the events to take place within the environment at the scripted time. The sequence of the $n-$back task was manipulated for each participant using the *Latin square technique* [5] to counterbalance the experiment.

   - *Dual−Experiment:* A 2 (tasks) $*$ 4 (levels i.e control, 0-back, 1-back, 2-back) within-subject experimental design was considered. In a within-subject design, every single participant is subjected to every single condition, including the control (CTRL) and has to perform every task included in the design. Participants performed the $n-$back task (zero, one, two) along with the DRT with their response to $n-$back being recorded to calculate accuracy. For the DRT, a stimulus i.e. a small electric vibrator (tractor) generating mild vibrations was attached to the participant's left or right arm depending on whether they are left-handed or right-handed. The response switch was attached to the index finger of the opposite hand.

Participants had to press the switch every time they felt a vibration, and also repeat out loud the number from $n-$back task at the same time. Participants were instructed that their primary task is to repeat numbers accurately while concentrating on the plus sign displayed on the screen in front of them and the secondary task is to press the response switch. Eye-tracking data were also recorded at the same time. Additionally, participants performed a control task with just the DRT alone. Subjective measure was recorded using the NASA task load index (NASA-TLX) form in which the participants had to rate each $n-$back task and control task on six different scales. NASA-TLX is a subjective, multidimensional assessment tool that rates perceived workload to assess a task [4].

- *Single−Experiment:* A 1 (task) $*$ 3 (levels) within-subject experimental design was considered. Participants performed the $n-$back task while concentrating on the plus sign on the computer screen in front of them without DRT. Eye-tracking data were recorded. Subjective measure was recorded using the NASA-TLX form in which the participants had to rate each $n-$back task on six different scales.

Due to a malfunction in the eye-tracking device, the data from participants 29 to 33 were excluded from further analysis. As a result, we consider only the data from the first 28 participants for the analysis presented in the remaining sections of this chapter.

2.3 Approach to Data Analysis

In the following sections, we present analysis of the collected data. Analysis is performed in two parts:

1. **Descriptive statistical Analysis:** Involves visual analysis of mean and standard deviation of the psychological data collected across 28 participants and

two stages (dual and single experiment).

2. **Inferential statistical Analysis:** Involves a statistical analysis of the mean of the psychological data collected across 28 participants and two stages (dual and single experiment) to validate observations from descriptive statistics.

   Descriptive statistics suggest that the psychological measures collected across 28 participants during the $n-$back task are significantly different depending upon the difficulty of the task. For instance descriptive statistics indicate that the mean of reaction time increase from $CTRL$ to $2-$back (Figure 2.4). As the sample size of the data under consideration is small (only 28 participants), results obtained from descriptive statistics might have occurred due to chance i.e. the psychological measures may not be different from each other depending on the task. To justify our findings i.e. psychological measures are different during the $n-$back task we analyze collected data using inferential statistics, i.e., through formal hypothesis testing.

   - *Following hypothesis were tested:*

   (a) **Hypothesis 1:** Mean reaction time increases with the increase in $n-$back difficulty.

   (b) **Hypothesis 2:** Mean of normalized pupil dilation collected during dual experiment increases with an increase in $n-$back difficulty.

   (c) **Hypothesis 3:** The means of NASA-TLX scales recorded during the dual experiment are significantly different. Mental demand, temporal demand, effort and frustration experienced by the participant's increases with $n-$back difficulty.

   (d) **Hypothesis 4:** Mean of normalized pupil dilation collected during single experiment increases with increase in $n-$back difficulty.

(e) **Hypothesis 5:** Mean of NASA-TLX scales recorded during single experiment are significantly different. Mental demand, temporal demand, effort and frustration experienced by the participant's increases with $n-$back difficulty.

(f) **Hypothesis 6:** The cognitive load experienced by an individual is higher during multitasking compared to a single task. Dual experiment had two tasks (i.e multitasking) compared to single experiment with one task. This hypothesis is tested by comparing mean of normalized pupil dilation from the dual experiment stage with mean of normalized pupil dilation from single experiment stage and the assumption is, pupil dilation is higher during the dual-experiment than the pupil dilation during single experiment.

(g) **Hypothesis 7:** Mean $n-$back accuracy recorded during the dual and single experiment is significantly different.

(h) **Hypothesis 8:** Mean mental demand scale recorded through NASA-TLX during dual and single experiment is significantly different.

- *Following procedure was followed for hypothesis testing:*

(a) **Testing the assumption of sphericity:** Sphericity is the condition where the variances of the differences between all combinations of related groups (levels) are equal. Sphericity is tested to determine the type of distribution of data under analysis. Violation of sphericity is when the variances of the differences between all combinations of related groups are not equal. Mauchly's test of sphericity is a formal way of testing the assumption of sphericity. Mauchly's test of sphericity tests the null hypothesis that the variances of the differences are equal. Thus, if Mauchly's test of sphericity is statistically significant ($p < .05$), we can reject the null hypothesis and accept the alternative hypothesis that the variances of the differences are not equal i.e., sphericity has been violated. If sphericity is not violated t-test or analysis of variance (ANOVA) is used to analyze the data. If sphericity is violated we use either Chi-square or Fisher test to analyze the data. As our data did not violate the condition of sphericity we have used ANOVA. The t-test is not used for analysis as it takes into consideration only two groups at a time whereas we have four groups. (Mauchly's test of sphericity is conducted only before repeated measures ANOVA) [22].

(b) **Repeated measure ANOVA:** A repeated-measures ANOVA is also referred to as a within-subjects ANOVA, which is a test to detect any overall differences between related means of different groups. ANOVA is used to investigate either changes in mean scores over three or more time points, or differences in mean scores under three or more different conditions. The repeated measures ANOVA tests two hypotheses, the null hypothesis states that the related means of different groups are equal and the alternative hypothesis states that the related means of different groups are not equal (at least one mean is different from another mean). Thus, if ANOVA is statistically significant ($p < .05$), we can reject the null hypothesis and accept the alternative hypothesis [26].

(c) **Post HOC for repeated measure ANOVA:** ANOVA detects the overall differences between related means of different groups but it does not tell which specific group differed. To determine which pair of groups (e.g. CTRL and $2-$back) are significantly different from each other we conduct a multiple-comparison test on our data (i.e reaction time and pupil dilation). Multiple comparison test analyzes two hypotheses, the null hypothesis states that the related means of two groups (at a time) are equal and the alternative hypothesis states that the related means of two groups (at a time) are not equal. Thus, if multiple comparison test is statistically significant ($p < .05$), we can reject the null hypothesis and accept the alternative hypothesis [21].

(d) **Effect size:** Effect size (*cohen's d*) is a statistical concept that measures the strength of the relationship between two variables on a numeric scale. The effect size was calculated using Cohen's d. Cohen's d is known as the difference of two related groups means and it is divided by the standard deviation (of the difference) from the data.

2.4 Data Analysis of the Dual Experiment

2.4.1 DRT Response Time

Descriptive Analysis of DRT Response Time

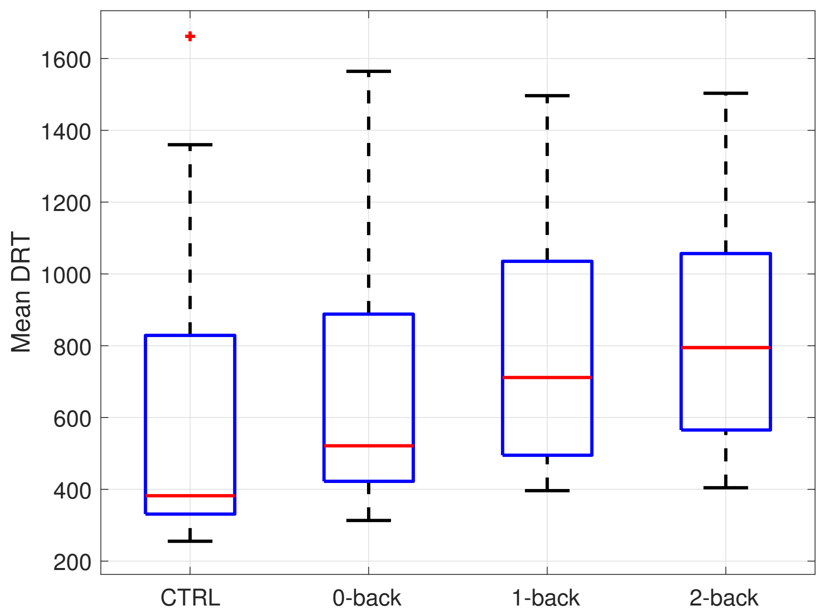
Figure 2.4 shows the characteristics of the mean reaction time with different difficulty levels in the $n-$back task. This figure is based on the data from 28 participants for each difficulty level; during the control task (denoted as CTRL) the participant is required to respond to the DRT vibrations without needing to respond to the $n-$back task. The reaction time increases with the difficulty level from CTRL until $2-$back. Reaction time below 100 ms (also known as a premature response) and above 2500 ms (also known as a un-requested response) were removed during data analysis. Also, the number of misses, where the participant did not press the response button, was not included.

Inferential Analysis of DRT Response Time

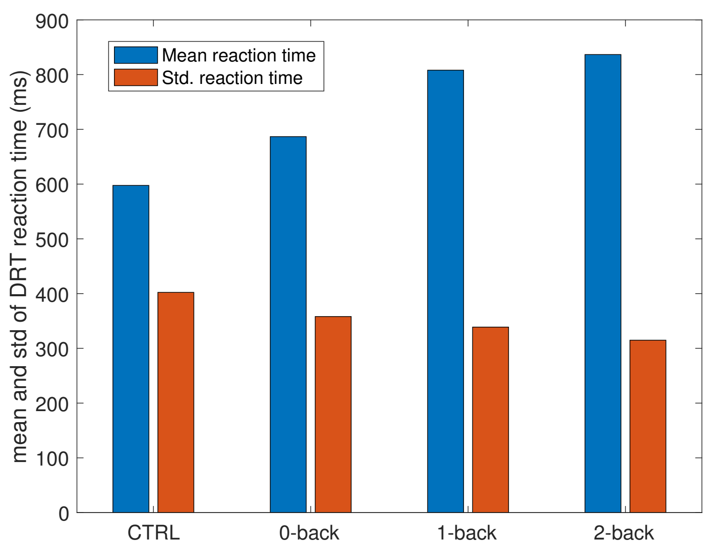
- **Hypothesis 1:** Mean reaction time increases with the increase in $n-$back difficulty.

Dual Experiment with two tasks (DRT and $n-$back) $*$ four levels (CTRL, $0-$back, $1-$back, $2-$back).

Mean reaction time data analysis using Mauchly's test [chi-square $x^2(5) = 7.85$, p =



(a) Box plot



(b) Mean and s.d. of the data in (a)

Figure 2.4: **Mean reaction time data of 28 participants.** The mean reaction time increases with difficulty i.e. shorter reaction times are observed for CTRL task and as the cognitive load increases reaction time also increases. Mean reaction time is highest during 2−back task.

.16] did not indicate any violation of sphericity.

Within subject or repeated measure ANOVA was considered to analyze mean response time data. The analysis revealed significant difference [F (3,78) = 3.016, $p < .034$, $cohen's\ d = .10$] between the overall mean of reaction time for different groups.

Post HOC analysis of the mean reaction time data revealed only two pair of groups i.e CTRL and 1−back [$p = .005$, $cohen's\ d = 1$] and, CTRL and 2−back [$p = .001$, $cohen's\ d = 1.26$] are significantly different from each other. Figure 2.5 shows that mean reaction time increases from CTRL task to 1−back and CTRL task to 2−back task i.e longer reaction time as the task got difficult.

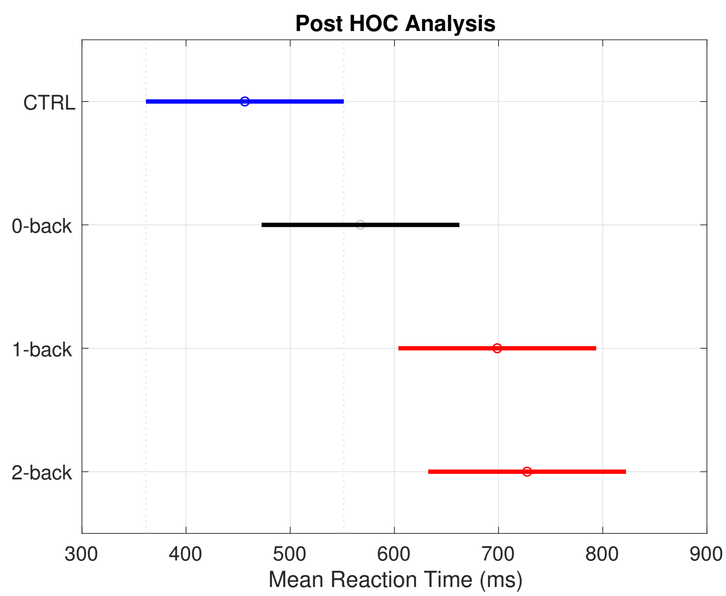


Figure 2.5: **Post HOC analysis for hypothesis 1.** Blue line indicates the task who's mean is significantly different than the other task (highlighted in red color). Black line represents the task who's mean does not differ significantly compared to the other tasks. 1−back and 2−back tasks have reaction time means significantly different from CTRL tasks. Longer reaction times are observed during 2−back and 1−back task as compared to CTRL task which indicate increasing in cognitive load with increase in task difficulty.

2.4.2 Missed DRT Trials

Descriptive Analysis of Total Number of Missed DRT Trials

The number of misses is defined as the number of times participants failed in responding to DRT vibrations [13]. The total number of missed DRT trials is calculated as the number of misses divided by trial number and multiplied by 100 to calculate the percentage of the missed DRT trials. Figure 2.6 represents the percentage mean of the total number of misses across 28 participants during four different tasks. The trial number for the individual task was approximately between 40-50 trials. The total number of misses is increasing from the CTRL task to 2−back task which indicates that cognitive load increased with task difficulty.

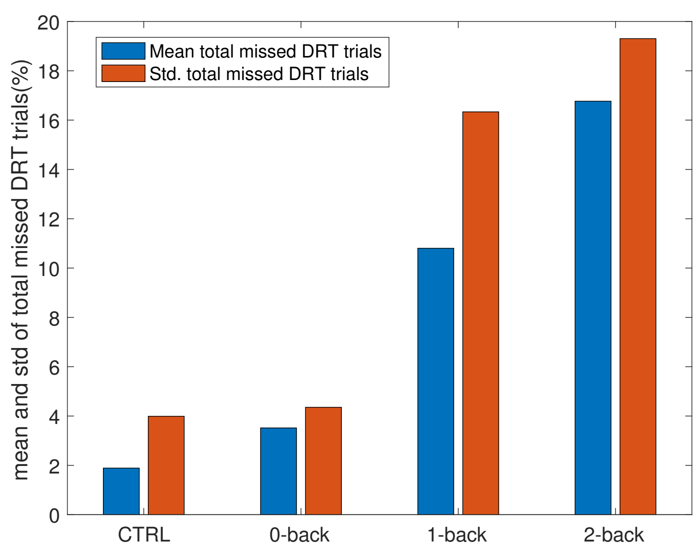


Figure 2.6: **Total number of missed DRT trials.** Total number of missed DRT trials increases with difficulty of the task.

2.4.3 Pupil Dilation

Descriptive Analysis of Pupil Dilation

Since the size of the pupil dilation might be different for each individual based on their physical characteristics, it is necessary to normalize the mean pupil dilation data. The pupil dilation data is normalized by dividing the measured pupil dilation by the average pupil dilation during the CTRL stage. Figure 2.7 shows the result of such normalization: the top plot shows the measured pupil dilation and the plot at the bottom shows the normalized pupil dilation.

Figure 2.8 Shows the plot of mean pupil dilation with different difficulty levels in the $n-$back task. This figure is based on the data from only 28 participants for each difficulty level. Pupil dilation data was collected when the participants performed the $n-$back task along with DRT. Pupil dilation increases with difficulty level from CTRL until $2-$back.

Inferential Analysis of Pupil Dilation

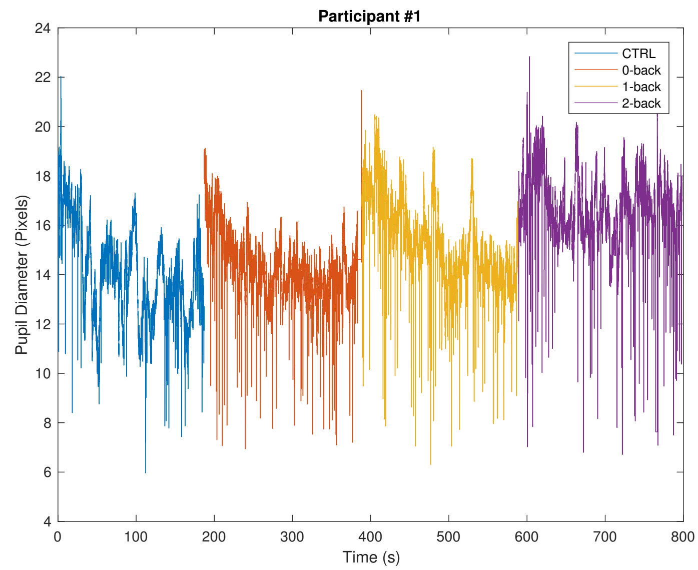
- **Hypothesis 2:** Mean of normalized pupil dilation collected during dual experiment increases with increase in $n-$back difficulty.

Dual Experiment with two tasks (DRT and $n-$back) x four levels (CTRL, $0-$back, $1-$back, $2-$back)

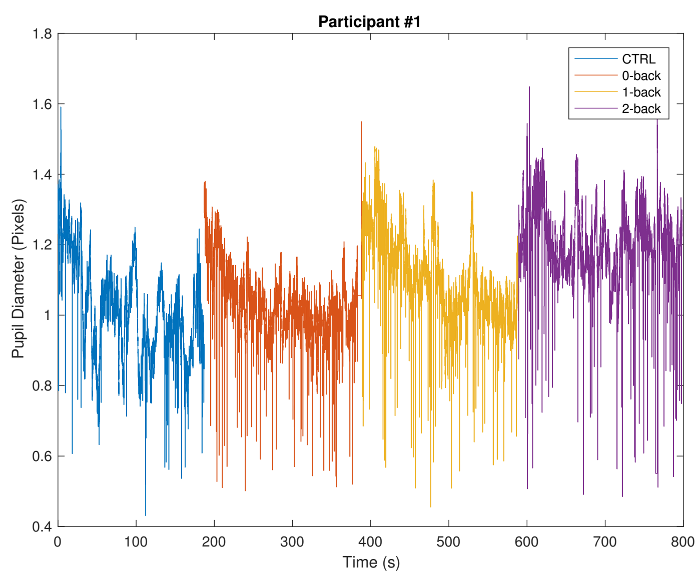
Mean of normalized pupil dilation data analysis using Mauchly's test [chi-square $x^2(5) = 12.117$, p = .7] did not indicate any violation of sphericity.

Within subject repeated measure ANOVA was considered to analyze pupil dilation data. The analysis revealed significant difference [F (3,78) = 4.578, p = .01, $cohen's$ d = 1.1] between the overall mean of normalized pupil dilation for different groups.

Post HOC analysis of the mean of normalized pupil dilation data revealed only three pair of groups i.e CTRL and $1-$back [p = .0081, $cohen's$ d = 0.8], CTRL and $2-$back [p = .00, $cohen's$ d = 1.2], and $0-$back and $2-$back [p = .00, $cohen's$ d = 1] are significantly different from each other. Figure 2.9 shows that mean of normalized pupil dilation increased from CTRL to $1-$back task and CTRL to $2-$back task i.e pupil dilation increases with the difficulty of the task. Pupil dilation results can be considered valid as the response time results show a similar pattern during CTRL to

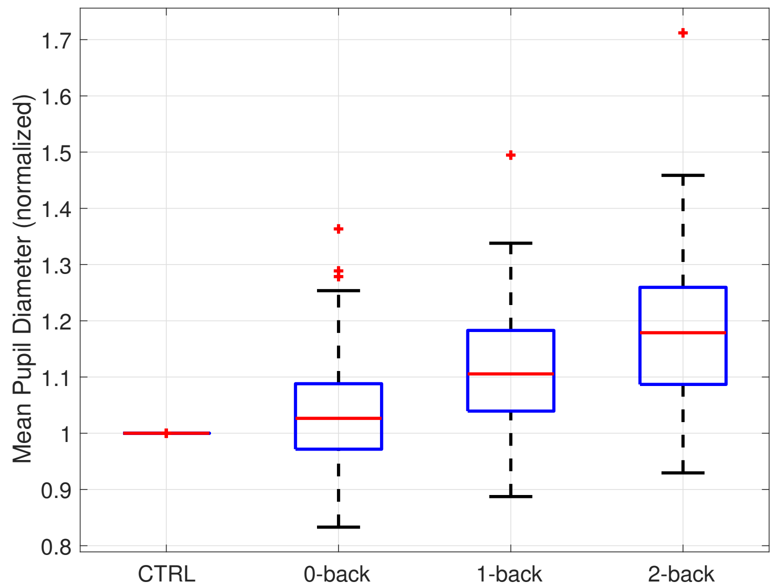


(a) Raw pupil dilation for single participant

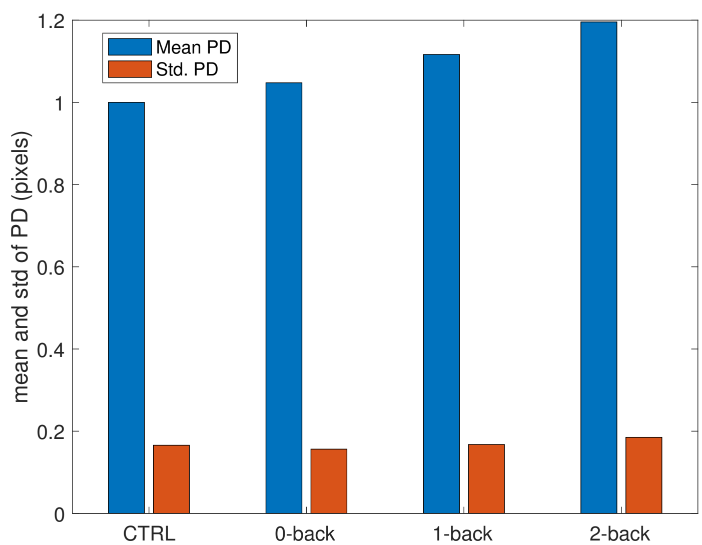


(b) Normalized pupil dilation for single participant

Figure 2.7: **Pupil dilation for single participant (dual experiment).** The pupil dilation data is normalized by dividing the measured pupil dilation by the average pupil dilation during the CTRL stage.



(a) Box plot



(b) Mean and s.d. corresponding to (a)

Figure 2.8: **Mean pupil dilation data of 28 participants (dual Experiment).** The mean pupil dilation increases with difficulty i.e. smaller pupil dilation is observed for CTRL task and as the cognitive load increases pupil dilation also increases. Pupil dilation is maximum during 2−back task.

1−back and CTRL to 2−back condition i.e an increased cognitive load experienced by the subjects.

2.4.4 NASA-TLX

Descriptive Analysis of NASA-TLX Data

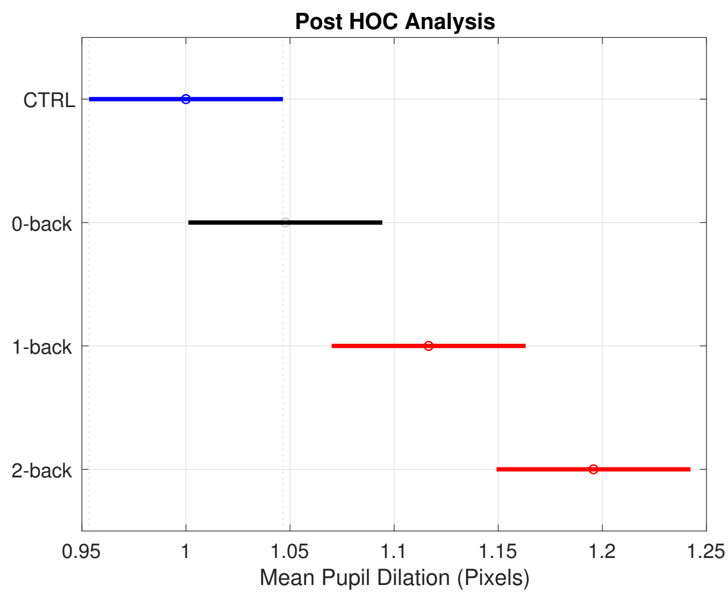
NASA-TLX data was collected after CTRL and each $n-$back task. Participants had to rate tasks performed on six different scales i.e. mental demand, physical demand, temporal demand, performance, effort, and frustration. For visualization and analysis only mental demand, temporal demand, effort and frustration scales were considered as the participant did not perform any task that was physically demanding and also the performance on the task was determined by calculating the accuracy.

Figure 2.10 (a) shows the plot for NASA-TLX scales i.e mental demand, temporal demand, effort and frustration across 28 participants and 2 (tasks) * 4 (levels). Mental demand, temporal demand, effort, and frustration increased with the difficulty of the task.

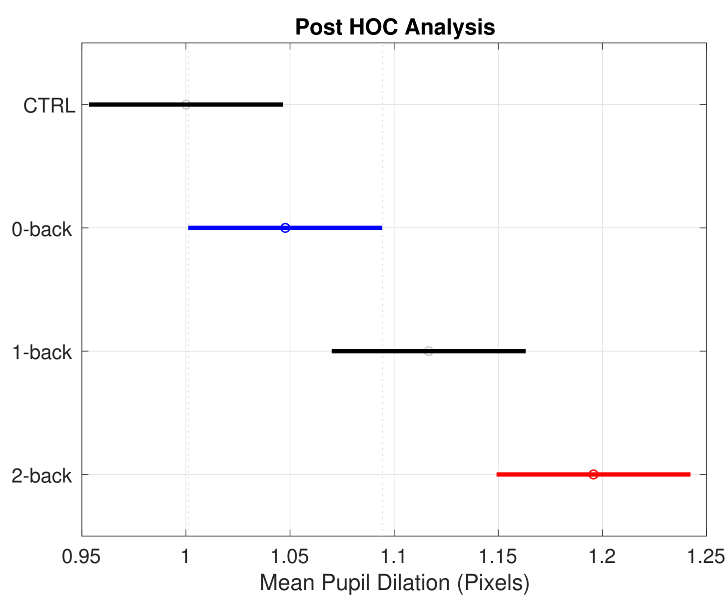
Inferential Analysis of NASA-TLX Data

- **Hypothesis 3:** Mean of NASA-TLX scales recorded during dual experiment are significantly different. Mental demand, temporal demand, effort and frustration experienced by the participant's increases with $n-$back difficulty.

A within-subject multivariate ANOVA with tasks (DRT and $n-$back with four conditions) as the independent variable and the four scales of NASA-TLX (mental demand, temporal demand, effort and frustration) as the dependent variables were considered. Significant effects of task were found for mental demand [F(3,23) = 4.52, p = .002,$cohen's\ d$ = 1.8], temporal demand [F(3,23) = 4.07, p = .01, $cohen's\ d$ = 1.6], and frustration [F(3,23) = 5.79, p = .004, $cohen's\ d$ = 1.68]. Ratings for effort did not differ across the task conditions.

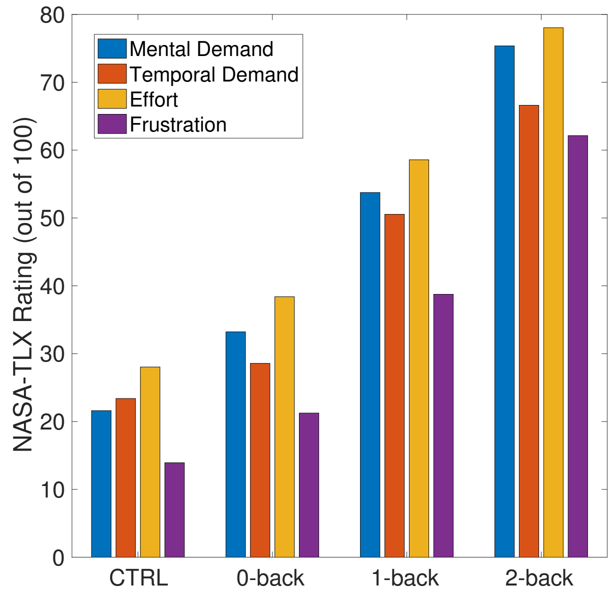


(a)

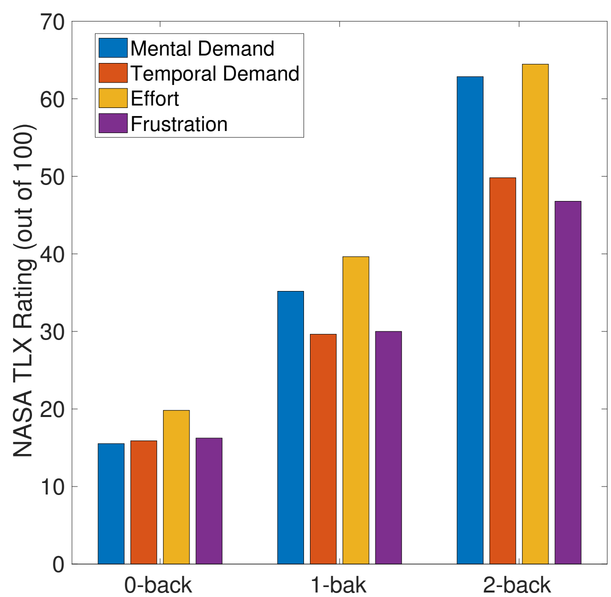


(b)

Figure 2.9: **Post HOC analysis for hypothesis 2.** Blue line indicates the task who's mean is significantly different than the other task (highlighted in red color). Black line represents the task who's mean does not differ significantly compared to the other tasks. The pupil dilation means of 1−back task and 2−back task are significantly different from CTRL task and pupil dilation mean of 0back is significantly different from 2−back. Dilation of increases with difficulty of the task i.e pupil dilates more during 1−back and 2−back tasks as compared to CTRL task which indicate increasing in cognitive load with increase in task difficulty.



(a) Mean of NASA-TLX rating from dual experiment



(b) Mean of NASA-TLX rating from single experiment

Figure 2.10: **Mean of NASA-TLX rating (dual and single experiment) across 28 participants.** Ratings of tasks performed during dual and single experiment are visualized using the mental demand, temporal demand, effort and frustration scales.

2.5 Data Analysis of the Single Experiment

In this subsection, we summarize the data visualization for single experiments where the DRT was not present. Without DRT, the measurements to be visualized are the pupil dilation and NASA-TLX data.

2.5.1 Pupil Dilation

Descriptive Analysis of Pupil Dilation

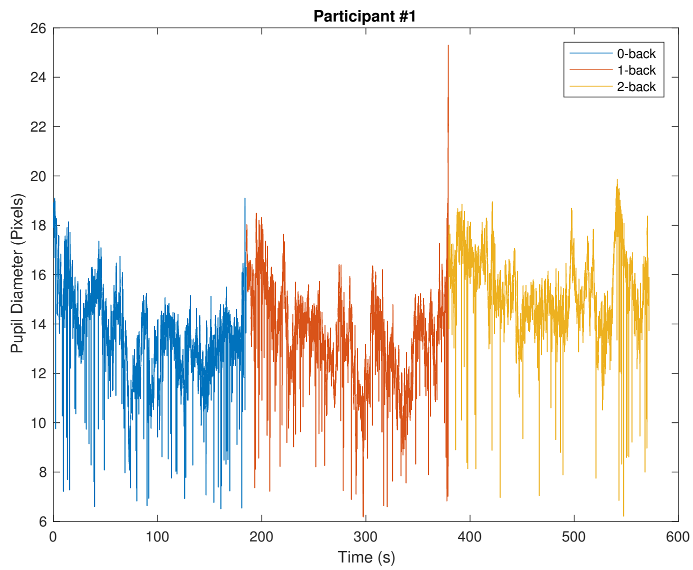
Similar to before, the pupil dilation information needs to be normalized since its initial size depends on individuals. The pupil dilation data for the single experiment is normalized by dividing the measured pupil dilation by the average pupil dilation during the 0-back stage.

Figure 2.12 shows the plot of mean of normalized pupil dilation with different difficulty levels in the $n-$back task. This figure is based on the data from only 28 participants for each difficulty level. Pupil dilation data were collected when the participants performed the $n-$back task without DRT. There was no control stage during this experiment. Pupil dilation increases with difficulty level from $0-$back until $2-$back.

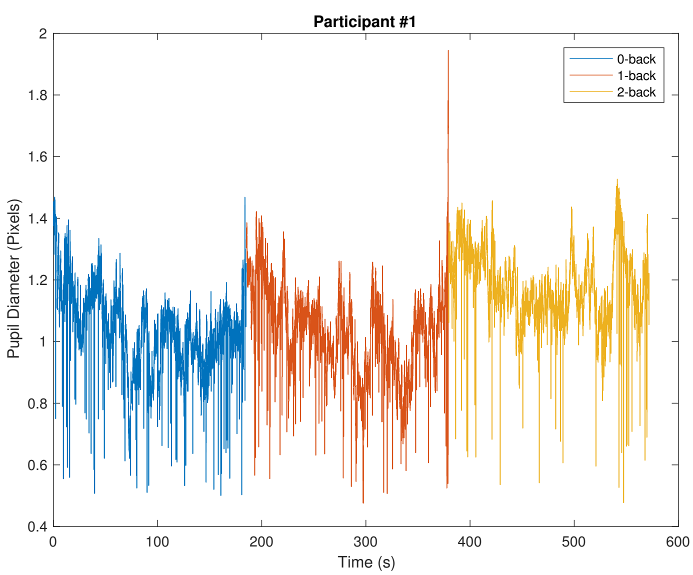
Inferential Analysis of Pupil Dilation

- **Hypothesis 4:** Mean of normalized pupil dilation collected during single experiment increases with increase in $n-$back difficulty.

Single Experiment with one tasks ($n-$back) x three levels ($0-$back, $1-$back, $2-$back) Mean of normalized pupil dilation data analysis using Mauchly's test [chi-square $x^2(3)$ = 8.1 , p = .8] did not indicate any violation of sphericity.
Within subject repeated measure ANOVA was considered to analyze pupil dilation data. The analysis revealed significant difference [F (2,52) = 4.578, p = .004, $cohen's$ d = 1.1] between the overall mean of normalized pupil dilation for different groups.

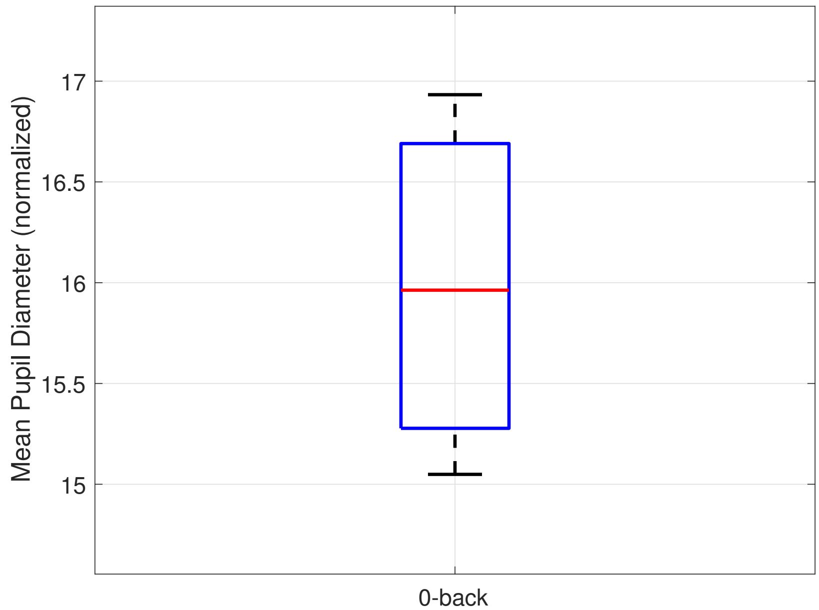


(a) Raw pupil dilation for single participant

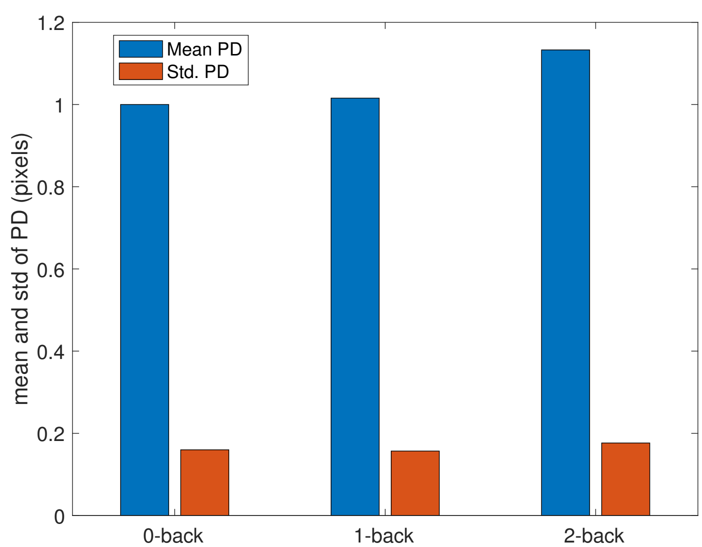


(b) Normalized pupil dilation for single participant

Figure 2.11: **Pupil dilation for single participant (single experiment).** The pupil dilation data is normalized by dividing the measured pupil dilation by the average pupil dilation during the CTRL stage.



(a) Box plot



(b) Mean and s.d. corresponding to (a)

Figure 2.12: **Mean pupil dilation data of 28 participants (single experiment).** The mean pupil dilation increases with difficulty i.e. smaller pupil dilation is observed for 0−back task and as the cognitive load increases pupil dilation also increases. Pupil dilation is maximum during 2−back task.

Post HOC analysis of the mean of normalized pupil dilation data revealed only two pair of groups i.e 0−back and 2−back [p = .00, $cohen's\ d$ = 1], and 1−back and 2−back [p = .00, $cohen's\ d$ = .9] are significantly different from each other. Figure 2.13 shows that mean of normalized pupil dilation increased from 0−back to 2−back task and 1−back to 2−back task i.e pupil dilation increases with the difficulty of the task.

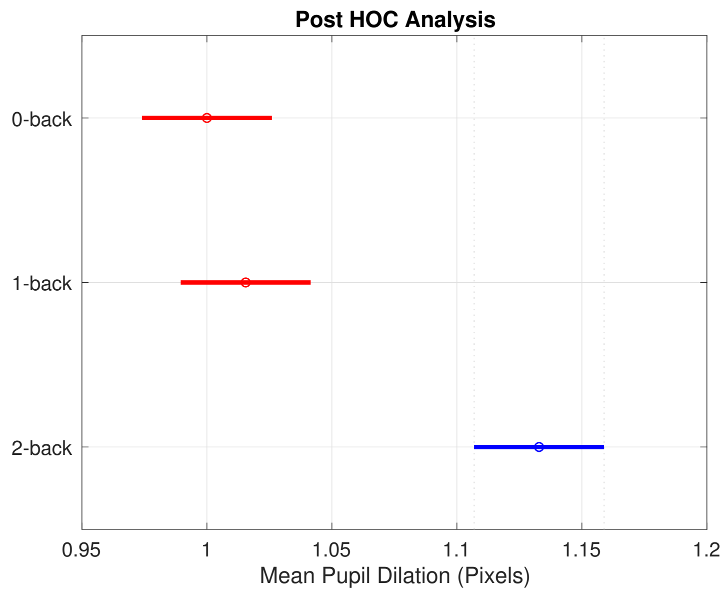


Figure 2.13: **Post HOC analysis of hypothesis 4.** The pupil dilation means of 0−back task and 1−back task are significantly different from 2−back task. Pupil dilation increases with difficulty of the task i.e pupil dilates more during 2−back tasks as compared to 0−back and 1−back task which indicate increasing in cognitive load with increase in task difficulty.

2.5.2 NASA-TLX

Descriptive Analysis of NASA-TLX Data

NASA-TLX data was collected after CTRL and each n−back task. Participants had to rate tasks performed on six different scales i.e. mental demand, physical demand, temporal demand, performance, effort, and frustration. For visualization and analysis only mental demand, temporal demand, effort and frustration scales were considered as the participant did not perform any task that was physically demanding and also the performance on the task was determined by calculating the accuracy.

Figure 2.10 (b) shows the plot for NASA-TLX scales i.e. mental demand, temporal demand, effort and frustration across 28 participants and 1 (task) * 3 (levels). Mental demand, temporal demand, effort, and frustration increased with the difficulty of the task.

Inferential Analysis of NASA-TLX Data

- **Hypothesis 5:** Mean of NASA-TLX scales recorded during single experiment are significantly different. Mental demand, temporal demand, effort and frustration experienced by the participant's increases with $n-$back difficulty.

A within-subject multivariate ANOVA with tasks ($n-$back with three conditions) as the independent variable and the four scales of NASA-TLX (mental demand, temporal demand, effort and frustration) as the dependent variables were considered. Significant effects of task were found for mental demand [F(2,24) = 17.41, p = .00, $cohen's\ d$ = 2.4], temporal demand [F(2,24) = 7.59, p = .002, $cohen's\ d$ = 1.9]. Ratings for effort and frustration did not differ across the task conditions.

2.6 Comparative Analysis of Dual and Single Experiment Data

In this section we compare data from dual and single experiment.

2.6.1 Pupil Dilation

Inferential Analysis of Pupil Dilation

- **Hypothesis 6:** Cognitive load experienced by an individual is higher during multitasking compared to single task.

This hypothesis is tested by comparing mean pupil dilation from the dual experiment stage with mean pupil dilation from single experiment stage and the assumption is mean pupil dilation is higher during the dual-experiment than the mean pupil dilation during single experiment. One-way ANOVA was considered to analyze the difference between mean pupil dilation data. The analysis revealed significant difference [F (1,167) $= 12.65$, $p = .0005$, $cohen's\ d = 0.5$] between the mean pupil dilation from dual experiment stage and single experiment stage. As expected multi-tasking imposes higher cognitive load on the subjects as compared to performing just the primary task (Figure 2.14) pupil dilation can successfully differentiate cognitive load depending upon the type of tasks.

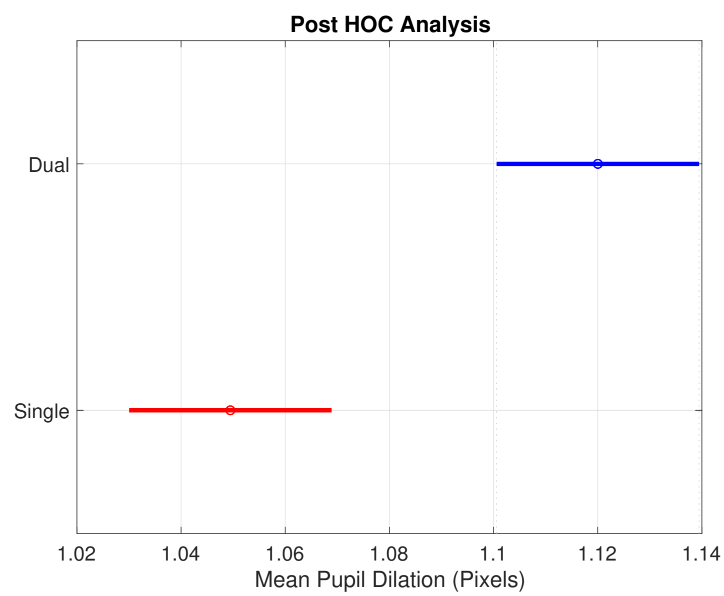


Figure 2.14: **Post HOC analysis of hypothesis 6.** Pupil dilation means are significantly different for dual and single experiment. Pupil dilates more during dual experiment as compared to single experiment.

2.6.2 $n-$back Accuracy

Descriptive Analysis of $n-$back Accuracy

Figure 2.15 shows the plot of $n-$back accuracy from the dual and single experiment stage.

Accuracy was calculated by comparing the original $n-$back series the participant heard and their response that was recorded in the form of audio for each $n-$back task. The first graph is the plot of overall-mean accuracy across 28 participants during zero, one and two back condition. The second graph is the plot of overall-s.d. of accuracy across 28 participants during zero, one and two back condition. The orange line in both the graphs represents the accuracy during the single experiment and the blue line indicates accuracy during the dual experiment. Figure 2.15 clearly represents that the mean accuracy reduces as the task becomes difficult for both dual as well as single experiment stage, but the accuracy during the dual experiment is lower than the accuracy during single experiment due to the fact that participants were performing two tasks i.e. $n-$back and DRT during dual experiment stage whereas participants were performing just the $n-$back task during the single experiment.

Inferential Analysis of $n-$back Accuracy

- **Hypothesis 7:** Mean $n-$back accuracy recorded during the dual and single experiment is significantly different.

This hypothesis is tested by comparing mean $n-$back accuracy from the dual experiment stage with mean $n-$back accuracy from single experiment stage and the assumption is that mean $n-$back accuracy is higher during the single experiment stage as compared to dual experiment stage. One-way ANOVA was considered to analyze the difference between the mean $n-$back accuracy data. The analysis revealed significant difference [F (1,167) = 5.6, p = .01, $cohen's\ d$ = 0.3] between the mean $n-$back accuracy from dual experiment stage and single experiment stage. As expected multi-tasking imposes higher cognitive load on the subjects as compared to

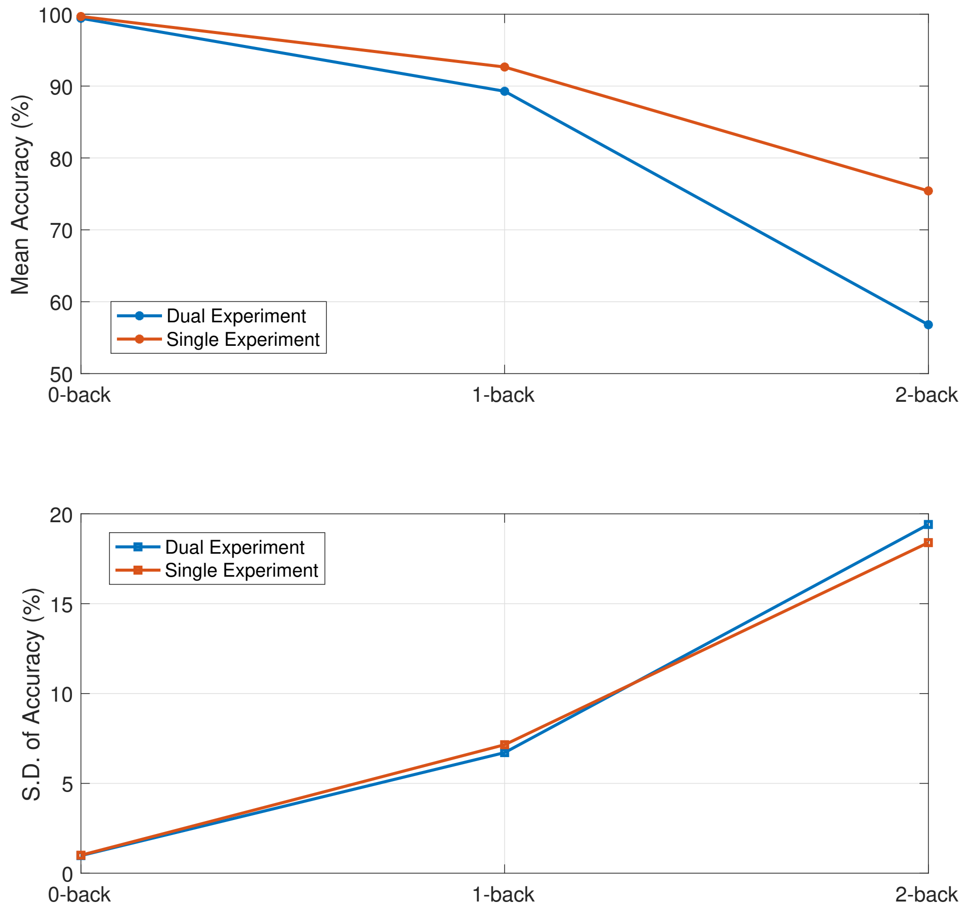


Figure 2.15: **Mean accuracy data of 28 participants (dual and single experiment).** Orange line in both the graphs represents the mean accuracy during single experiment and blue line indicates mean accuracy during dual experiment.

performing just the primary task, which is reflected by the low accuracy during dual experiment. Mean $n-$back accuracy is high during the single experiment as compared to the dual single experiment (Figure 2.16).

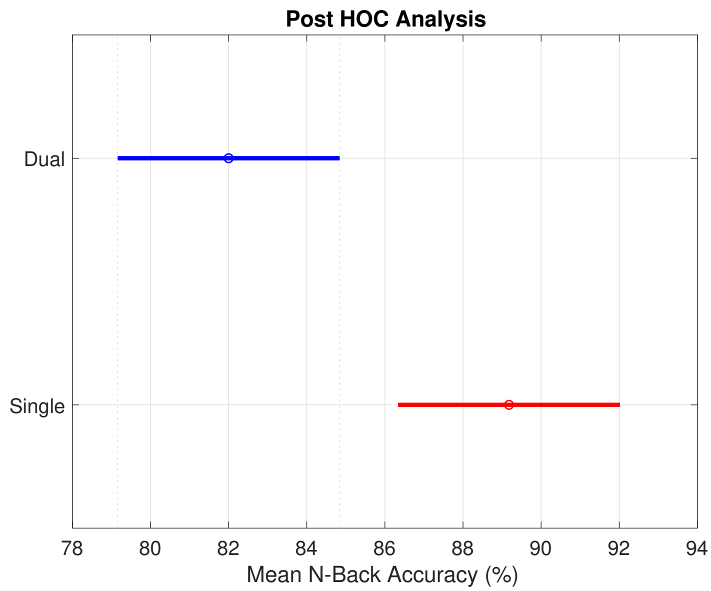


Figure 2.16: **Post HOC analysis of hypothesis 7.** $n-$back accuracy means are significantly different for dual and single experiment. Accuracy is higher during single experiment as compared to dual experiment.

2.6.3 NASA-TLX

Inferential Analysis of NASA-TLX data

- **Hypothesis 8:** Mean mental demand scale recorded through NASA-TLX during dual and single experiment is significantly different.

This hypothesis is tested by comparing mean mental demand from the dual experiment stage with mean mental demand from single experiment stage and the assumption is that mean mental demand is higher during the dual experiment as compared to the single experiment. One-way ANOVA was considered to analyze the difference between mean mental demand data. The analysis revealed significant difference [F

(1,161) = 6.9, $p = .0093$, $cohen's\ d = .5$] between the mean mental demand from dual experiment stage and single experiment stage. As expected multi-tasking imposes higher cognitive load on the subjects as compared to performing just the primary task, which is reflected by the higher mental demand dual experiment. (Figure 2.16) mean mental demand is higher during the dual experiment as compared to the single experiment.

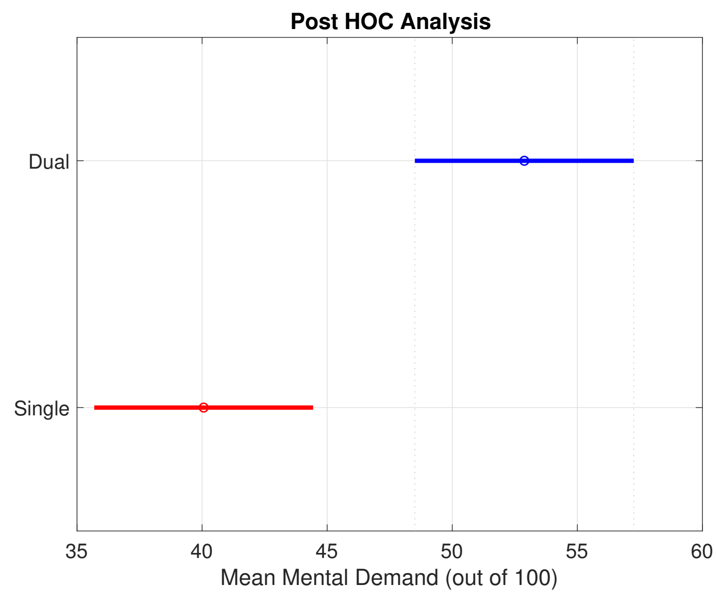


Figure 2.17: **Post HOC analysis of hypothesis 8.** Mental demand means are significantly different for dual and single experiment. Mental demand is higher during the dual experiment as compared to single experiment.

2.7 Conclusions and Discussions

In this chapter, we investigated the feasibility of using pupil dilation as a measure of cognitive load in advanced driver assistance systems (ADAS); as such, a low cost eye-tracker used to measure the pupil dilation without imposing any physical restrictions on the participants.

The experiments were divided in to *dual* and *single* ones. Furing the dual experiment, the participants performed the detection response task (DRT) in addition to performing the $n-$back memory task that required mental work that increased with the value of n; in addition to other measurements, the reaction time (RT) and the pupil dilation (PD) were used in the subsequent analysis. During the single experiments, the participants performed the $n-$back memory task only; here, the measured PD was used for analysis.

The dual experiments consisted of the following four stages in increasing order of cognitive load: CTRL (which stands for control), 0−back, 1−back and 2−back. Statistical inference analysis of recorded data (RT and PD) resulted in the conclusions shown in Table 2.4 about the statistical differences of the measured pairs.

Table 2.4: **Result of statistical difference (dual experiment)**

	RT	**PD**
[CTRL] to [0−back]	False	False
[CTRL] to [1−back]	True	True
[CTRL] to [2−back]	True	True
[0−back] to [1−back]	False	False
[0−back] to [2−back]	False	True
[1−back] to [2−back]	False	False

The single experiments consisted of the following three stages in increasing order of cognitive load: 0−back, 1−back and 2−back. There was no CTRL stage. Statistical inference analysis of recorded data (PD) resulted in the conclusions shown in Table 2.5 about the statistical differences of the measured pairs.

Table 2.5: **Result of statistical difference (single experiment)**

	PD
[0−back] to [1−back]	False
[0−back] to [2−back]	True
[1−back] to [2−back]	True

Findings of the pupil dilation coincide with DRT findings for the task CTRL to 1$-$back and CTRL to 2$-$back which helps support our objective i.e non-invasive or remote eye tracking (measuring pupil dilation) is a viable solution for detecting cognitive load experienced by an individual. Additionally, findings of pupil dilation can be validated from the analysis of NASA-TLX scales which indicate that mental demand and frustration imposed by the task on the participant increases with task difficulty.

Comparison of the pupil dilation data from the dual and single experiment shows that cognitive load experienced during two simultaneous tasks is higher than cognitive load imposed by a single task; this is evident from the higher pupil dilation during the dual experiment stage. In other words DRT acts as an additional cognitive load during the dual experiment. Further this finding was confirmed by the comparison of mental demand scale (NASA-TLX) from dual and single experiment, which revealed that mental demand is higher during the dual experiment as compared to single experiment; additionally, $n-$back accuracy comparison showed lower accuracy during dual experiment as compared to single experiment. Pupil dilation can evidently detect cognitive load experienced by an individual; besides, it can also detect additional cognitive load imposed by a second task.

This study has one limitation: participants were not driving even though this study is dedicated to improving the ADAS in semi-autonomous vehicles. However, it must be pointed out that many similar studies conducted in a controlled laboratory setting have provided findings that were then replicated in a more realistic environment [12]. Further, cognitive load detection based on pupil dilation has other applications in human-machine system automation [19, 20] where the findings of this chapter will be useful.

Even though the pupil dilation showed comparable performance to DRT as a detector of cognitive load, there are known limitations of PD that is not tested in this chapter. For example, the pupil dilation is affected by other stimuli, such as external light, whereas the DRT does not suffer from such external factors. Further research is required for repeating the same finding in a simulated as well as real driving environment is crucial to developing this towards real-world applications. Based on the data reported in this chapter, further studies will be needed to investigate the following three aspects:

1. *Applying signal processing to pupil dilation data:* Developing signal processing approach for improving the detectability of cognitive load through pupil dilation measurements. Raw pupil dilation data is masked by noise which reduces the capability of data to precisely predict cognitive load. Signal processing of the data can help detect cognitive load accurately for multiple difficulty levels.

2. *Data fusion method:* In addition to pupil dilation, there are other information that can be potentially combined for improved cognitive load detection.

3. *Predictive model:* Developing a machine learning model to accurately predict cognitive load through pupil dilation. Such a model will exploit the "training data" collected offline to train as such it can be used on new subjects (and applications) without the requirement for the lengthy training phase.

2.8 Bibliography

[1] "Delayed digit recall (n-back) task." [Online]. Available: http://agelab.mit.edu/delayed-digit-recall-n-back-task pages 15, 52

[2] "DRT detection response task website," https://www.redscientific.com/detection-response-task.html, accessed: 2018-20-11. pages 14

[3] "gazept gazepoint eye tracker website," https://www.gazept.com/, accessed: 2018-20-11. pages xii, 13

[4] "Nasatlx." [Online]. Available: https://humansystems.arc.nasa.gov/groups/TLX/ pages 18

[5] *Latin Square Designs.* New York, NY: Springer New York, 2008, pp. 297–297. [Online]. Available: https://doi.org/10.1007/978-0-387-32833-1_223 pages 17

[6] "European commission - traffic safety synthesis 2016," Feb 2017. [Online]. Available: https://www.nrso.ntua.gr/european-commission-traffic-safety-synthesis-2016/ pages 4, 8

[7] J. Beatty and D. Kahneman, "Pupillary changes in two memory tasks," *Psychonomic Science*, vol. 5, no. 10, pp. 371–372, 1966. pages 10

[8] J. Beatty, B. Lucero-Wagoner *et al.*, "The pupillary system," *Handbook of psychophysiology*, vol. 2, no. 142-162, 2000. pages 10

[9] A. Beck and K. J. Kek, "Pupillometry and sensor fusion for monitoring and predicting a vehicle operator's condition," Feb. 7 2019, uS Patent App. 15/666,017. pages 9

[10] E. Bekiaris and A. Stevens, "Common risk assessment methodology for advanced driver assistance systems," *Transport reviews*, vol. 25, no. 3, pp. 283–292, 2005. pages 9

[11] F. Biondi, M. Leo, M. Gastaldi, R. Rossi, and C. Mulatti, "How to drive drivers nuts: Effect of auditory, vibrotactile, and multimodal warnings on perceived urgency, annoyance, and acceptability," *Transportation Research Record*, vol. 2663, no. 1, pp. 34–39, 2017. pages 9

[12] O. E Kang, K. E Huffer, and T. Wheatley, "Pupil dilation dynamics track attention to high-level information," *PloS one*, vol. 9, p. e102463, 08 2014. pages 43

[13] I. O. for Standardization, "Road vehiclestransport information and control systemsdetection–response task (drt) for assessing attentional effects of cognitive load in driving." 2016. pages 5, 10, 14, 25, 49, 52

[14] S. G. Hart, "Nasa-task load index (nasa-tlx); 20 years later," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, no. 9. Sage publications Sage CA: Los Angeles, CA, 2006, pp. 904–908. pages 11

[15] T. Hong and H. Qin, "Drivers drowsiness detection in embedded system," in *2007 IEEE International Conference on Vehicular Electronics and Safety*. IEEE, 2007, pp. 1–5. pages 9

[16] M. Kuehn, T. Hummel, and J. Bende, "Benefit estimation of advanced driver assistance systems for cars derived from real-life accidents," in *21st International Technical Conference on the Enhanced Safety of Vehicles ESV*, vol. 15, 2009, p. 18. pages 8

[17] N. Lavie, "Attention, distraction, and cognitive control under load," *Current directions in psychological science*, vol. 19, no. 3, pp. 143–148, 2010. pages 9

[18] P. Mannaru, B. Balasingam, K. Pattipati, C. Sibley, and J. Coyne, "Cognitive context detection for adaptive automation," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 60, no. 1. Sage Publications Sage CA: Los Angeles, CA, 2016, pp. 223–227. pages 10

[19] ——, "Cognitive context detection using pupillary measurements," in *Next-Generation Analyst IV*, vol. 9851. International Society for Optics and Photonics, 2016, p. 98510Q. pages 43

[20] P. Mannaru, B. Balasingam, K. Pattipati, C. Sibley, and J. T. Coyne, "Performance evaluation of the gazepoint gp3 eye tracking device based on pupil dilation," in *International Conference on Augmented Cognition*. Springer, 2017, pp. 166–175. pages 43

[21] G. A. Milliken and D. E. Johnson, *Analysis of messy data, volume II: Nonreplicated experiments*. Chapman and Hall/CRC, 2017. pages 22

[22] B. Nagarsenker and K. Pillai, "The distribution of the sphericity test criterion," *Journal of Multivariate Analysis*, vol. 3, no. 2, pp. 226–235, 1973. pages 21

[23] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. Van Gerven, "Cognitive load measurement as a means to advance cognitive load theory," *Educational psychologist*, vol. 38, no. 1, pp. 63–71, 2003. pages 9

[24] S. Petridis, T. Giannakopoulos, and C. D. Spyropoulos, "Unobtrusive low cost pupil size measurements using web cameras," *arXiv preprint arXiv:1311.7327*, 2013. pages 11

[25] A. Shaout, D. Colella, and S. Awad, "Advanced driver assistance systems-past, present and future," in *International Computer Engineering Conference*, 2011, pp. 72–82. pages 4, 5, 8

[26] C. N. von Ende, "Repeated-measures analysis," *Design and analysis of ecological experiments*, vol. 8, pp. 134–157, 2001. pages 21

[27] B. Xie and G. Salvendy, "Prediction of mental workload in single and multiple tasks environments," *International journal of cognitive ergonomics*, vol. 4, no. 3, pp. 213–242, 2000. pages 9

[28] R. A. Young, L. Hsieh, and S. Seaman, "The tactile detection response task: preliminary validation for measuring the attentional effects of cognitive load," 2013. pages 10, 49

Chapter 3

Comparison of Cognitive Load Classification Based On Pupil Dilation and DRT Reaction Time

3.1 Introduction

Pupil dilation and response time are considered as reliable parameters for indexing information processing loads [9, 17]. Time locked averaging of pupil dilation data concerning events provoking cognitive activity can be related to changes in the central nervous system that are systematically associated with cognitive processing i.e. when people are faced with a challenging cognitive task, their pupil dilates. This phenomenon is also called task-evoked pupillary response [4]. Change in pupil dilation due to the difficulty of the task is relatively small but have a lot of predictive strength in terms of cognitive load detection.

Response time collected through the detection response task is an accepted standard for capturing the attentional aspect of cognitive load [5]. Response time varies depending upon the cognitive demand placed by the task. Longer reaction times are observed when a task has high cognitive demand [21]. With the development in fields of eye-tracking and devices measuring response time, cognitive load detection has become a little easier [16]. The big question to answer here is how can the cognitive load be classified among different difficulty levels. Cognitive load varies depending upon the difficulty of the task and the number of mental resources available to complete the task [19]. It becomes crucial to classify cognitive load depending on the task to help individual in better resource allocation as well as improving the performance.

Cognitive load can be classified with the help of classification algorithms [15]. A classification algorithm trains on the available information such as pupil dilation with preassigned class labels and tries to classify the new dataset. Many different classifying algorithms can help in distinguishing between the cognitive load. A classifier with high accuracy can be used to detect the cognitive context in a variety of applications that require human-machine interaction. One such example is designing an adaptive advanced driver assisted systems also referred to as ADAS. ADAS generally assist the driver to ease the driving experience or to prevent any unusual circumstances. The problem with existing ADAS is that they are not completely adaptive according to the driver's mental state. A well trained cognitive load classifier based on pupil dilation and reaction time features can help develop such an adaptive system [10]. For instance, if a driver is experiencing cognitive load which is higher than normal, an adaptive ADAS system can be designed to measure and classify the cognitive load, making it adapt i.e. improvise further course of action by providing additional assist or displaying a warning message that can help prevent any unsafe condition. Cognitive load classification demonstrated in this chapter can be helpful in such environments.

Researchers have demonstrated the classification of cognitive load using psychological parameters similar to pupil dilation and reaction time. MH kutila, M jokela and T makinen [11] used support vector machine classier for classifying cognitive load using eye gaze, head and lane-keeping variance with approximately 60% to 80% confidence. Similar research was done by Bo yin and Natalie ruiz [20], they used a Gaussian mixture model to classify 3 levels cognitive load based on speech features with a classification accuracy of 71.1%.

Observations of this chapter are listed below:

1. *Classification of cognitive load using multiple features:* We have demonstrated the classification of cognitive load by combine two features i.e. pupil dilation and reaction time.

2. *Comparison of classification algorithms for cognitive load classification:* Various classification algorithms are available for classifying data, we test the accuracy of the three most commonly used algorithms i.e. support vector machine, logistic regression, and k-means to determine the best fit for classifying cognitive load in this research.

3. *Demonstrate cognitive load detection through the signal to noise ratio (SNR):* We compare the detectability in terms of a newly proposed metric called SNR [12].

4. *Comparison of single feature classification with classification using data fusion:* We compare the accuracy of classification obtained using the single feature (pupil dilation; reaction time) with classification accuracy obtained from data fusion to test the effectiveness of data fusion in cognitive load classification.

The rest of this chapter is organized as follows: Section 3.2 provides the description of the procedure followed for data collection, apparatus used, and tasks involved in the experiment; Section 3.3 describes three different classification algorithms used in this chapter.; Section 3.4 explains the data fusion approach, structure of the data, procedure followed for classification and accuracy analysis of the algorithms used. Section 3.5 illustrates the performance analysis of classifier with individual and fused features. and Section 3.6 concludes this chapter.

3.2 Data Collection

In this section we provide a short description of procedure followed for collecting pupil dilation and reaction time data which is used later in this chapter for classification analysis. The details of the data collection setup, procedures and graphical/statistical analysis of the collected data is presented in [3].

- Thirty-three participants were recruited.
- Reaction time data was collected using the detection response task referred to as DRT. DRT consisted of a stimulus in the form of vibrations generated by a small tactile generator and response was in the form of a microswitch. The stimulus was attached to the participant's forearm and the microswitch was attached to the index finger of the other hand. The participant had to respond by pressing the microswitch every time they feel a vibration produced by the stimulus. The time between the vibration generated and the participant pressed the microswitch was recorded as the reaction time [5].
- Pupil dilation data was collected using an eye-tracking device.
- Delayed digital recall task referred to as *n-back*, with three difficulty levels was used to permutate the cognitive difficulty. The n-back task has a serial presentation of a stimulus in the form of audio which is placed several seconds apart and the participants had to repeat out loud the number they just heard [1]. The three difficulty levels are:
    1. 0-back: Participants heard a series of number and they had to repeat out loud the same number they just heard.
    2. 1-back: Participants heard a series of numbers and they had to repeat out loud one number previous to the number they just heard.
    3. 2-back: Participants heard a series of numbers and they had to repeat out loud two numbers previous to the number they just heard.

- A within-subject design was used in order to carry out the experiment. Entire experiment was divided into two stages i.e. dual experiment and single experiment.

- During the dual experiment participant had to perform two tasks i.e. DRT and *n-back*. In addition to three $n-$back tasks participants also performed a control task (CTRL). During the CTRL task participants had to respond just to the DRT vibrations without having to repeat any numbers. Pupil dilation and response time were recorded during this stage.

- During the single experiment participant had to perform just the *n-back* task. Only pupil dilation data was recorded during this stage. There was no CTRL task during this stage. In this chapter, the data i.e pupil dilation and reaction time from the dual experiment is considered. The data used for analysis in the present chapter is summarized in Table 3.6 and Table 3.7

3.3 Classification algorithms

In this section, we provide a description of the classification algorithms used for classifying the cognitive load experienced by the participant using response time (RT) and pupil dilation (PD). Also, a recently developed method called SNR is used for detecting the cognitive load. SNR is discussed in detail in the latter part of this section.

For this chapter, the classification has been limited only to binary classification. As the classification model in this chapter is demonstrated in reference to the ADAS application. When developing a cognitive load adaptive ADAS the measured cognitive load, for instance through pupil dilation will make the operation of ADAS twofold i.e if the cognitive load is low a warning message will be displayed and if the cognitive load is high ADAS will be activated. This will be done by classifying the detected

Table 3.6: Mean reaction time (RT in milliseconds)

Participant ID	CTRL	0-back	1-back	2-back
1	697.46	670.43	598.07	636.23
2	321.56	366.79	466.31	539.48
3	389.72	381.10	503.96	554.10
4	306.63	607.00	808.06	629.52
5	330.74	520.37	476.27	429.56
6	418.10	626.07	946.48	765.88
7	543.08	629.20	666.94	576.40
8	1056.43	1175.31	1282.96	1503.26
9	255.39	313.25	396.32	431.00
10	365.40	415.77	407.76	529.28
11	374.63	415.77	632.36	628.35
12	1356.38	987.24	885.05	1082.15
13	339.89	486.25	729.35	1059.45
14	331.27	474.10	674.96	941.90
15	240.65	366.83	418.40	478.40
16	342.30	409.45	435.48	630.54
17	415.68	522.12	693.56	859.93
18	285.65	380.18	485.92	510.83
19	362.02	461.04	850.31	823.42
20	374.07	1401.69	1309.05	1000.23
21	827.10	920.43	1104.00	1159.50
22	830.45	855.67	1212.03	913.34
23	290.30	483.90	534.39	984.59
24	404.41	500.95	1223.31	404.41
25	260.72	319.44	418.98	727.33
26	505.58	554.72	681.10	708.10
27	279.26	364.81	411.94	555.83
28	275.76	278.92	312.49	304.25

Table 3.7: Mean pupil diameter (PD in pixels)

Participant ID	**CTRL**	**0-back**	**1-back**	**2-back**
1	1	1.01	1.08	1.17
2	1	0.98	1.03	1.27
3	1	1.10	1.19	1.33
4	1	1.06	1.05	1.02
5	1	1.28	1.17	1.14
6	1	0.99	1.06	1.32
7	1	0.83	0.89	0.95
8	1	1.36	1.49	1.39
9	1	1.10	1.12	1.18
10	1	0.99	1.02	1.71
11	1	0.96	1.16	0.98
12	1	1.03	1.20	1.24
13	1	0.91	1.10	1.09
14	1	1.07	1.04	1.25
15	1	1.02	1.16	1.13
16	1	1.12	1.13	1.17
17	1	1.01	1.15	1.19
18	1	0.95	0.97	1.17
19	1	0.92	1.11	1.03
20	1	1.25	1.23	1.25
21	1	0.92	0.89	0.93
22	1	1.03	1.04	1.14
23	1	1.00	1.34	1.24
24	1	1.08	1.05	1.08
25	1	1.04	1.05	1.02
26	1	0.95	0.99	1.40
27	1	1.06	1.22	1.21
28	1	1.29	1.32	1.46

cognitive load as either low or high. For this reason, the further sections will only discuss binary classification. The presented way of classification can be referred to as group learning. Group learning can be defined as learning a pattern from a group and generalizing it for the entire population. In this chapter mean pupil dilation and mean reaction time across 28 participants are analyzed and classified for detecting cognitive load.

3.3.1 Types of classification algorithms used

1. **Support vector machine:** Support vector machines (SVM) is a widely used classification technique. SVM separates two classes of data by finding the best hyperplane which separates the class1 data from class2 data. SVM selects the hyperplane which has the largest margin separating the two classes. The goal of SVM is to develop a model which predicts the target values given only the test data attributes [8].

   SVM is just like 1 layer or multi-layer neural networks. SVM works on a concept called support vectors i.e the data points that lie close to the decision surface (hyperplane) [7]. These points are difficult to classify, SVM finds the best hyperplane which separates these points. Linearly separable data sets are classified in a hyperplane and non-linearly separable data sets are classified using the kernel function. Kernels are usually used to classify non-linearly separable data sets by gaining linear separation. Linear separation is achieved by mapping the data to a higher-dimensional space. Some of the examples of kernel functions are polynomials, radial bias also known as Gaussian kernel function and multi-layer perceptron or sigmoid function [2]. For our analysis, we have used the gaussian kernel function. Matlab's inbuilt support vector machine model [14] was used to classify the reaction time and pupil dilation data. SVMModel from Matlab was feed with a matrix of input features (each row was one observation and each column was one feature), class labels corresponding to each value of the predictor data and Kernel function depending upon the type (linearly separable or non linearly separable) of input features.

2. **Logistic regression:** Logistic regression is a statistical method for predicting binary classes. A logistic function is used in order to predict the probability that particular data belongs to one of the class under consideration.

   Logistic regression is a binary and multi class classifier. Logistic regression works on a hypothesis also called as the logistic function (equation 3.1). Logistic regression uses a threshold for the hypothesis (equation 3.2). If the calculated hypothesis function is smaller than 0.5 logistics regression model predicts that the data belongs to class with label 0. If hypothesis function is greater than 0.5 logistics regression model predicts that the data belongs to class with label 1. In our case the label 0 and 1 corresponds to the class CTRL, 0, 1, and 2−back. In equation 3.1 x is the input feature feed to the logistic regression i.e. reaction time and pupil dilation. Predictions are made based on the input feature and θ. Initially θ is set to zero. One of the crucial steps in performing logistic regression is selecting the value of θ. Logistic regression uses a cost function in order to determine θ. Cost function is known as the penalty classifier pays if θ is large. In order to minimize cost, gradient descent is carried out. Gradient descent produces a value of theta which is minimum and optimal for the classifier thus reducing the penalty. After calculating the optimal theta classifier can accurately make class predictions for the input features.

$$h_{\mathbf{\Theta}}(x) = \frac{1}{1 + \exp^{\mathbf{\Theta}^{\intercal}} x} \tag{3.1}$$

$$\begin{aligned} &if\ h_{\mathbf{\Theta}}(x) \geq 0.5, predict\ y = 1 \\ &if\ h_{\mathbf{\Theta}}(x) \leq 0.5, predict\ y = 0 \end{aligned} \tag{3.2}$$

where:

$h_{\mathbf{\Theta}}(x)$ = Prediction

$\mathbf{\Theta}$ = Cost

x = Input feature

3. **K-means:** k-means allocates a specific location to every feature (response time and pupil dilation) in a space, then it locates the centers for the individually defined clusters or classes in a multidimensional space. Each point is then assigned to the cluster whose arbitrary mean vector is closest. The procedure continues until there is no significant change in the location of class mean vectors between successive iterations of the algorithms [18]. K-means is an iterative, data partitioning algorithm. K-means groups similar data points together in respective clusters. Cluster is a collection of data points combined together because of certain similarities. K-means works on the principle of centroid allocation. A centroid is the imaginary or real location representing the center of the cluster. K-means algorithm identifies the total clusters through the number of centroids and then allocates every data point to the nearest cluster. The means in the k-means refers to averaging of the data; that is, finding the centroid [6]. Matlab's inbuilt k-means algorithm was used in order to classify the cognitive load based on reaction time and pupil dilation [13].

   Matlab's k-mean function requires input in the form of features on which classification is based, the number of clusters and, distance metric. The distance metric determines the function to be used in order to calculate the distance of a particular point from the centroid. Matlab's *sqeuclidean* function was used as the distance metric. Sqeuclidean is defined as the squared of the distance between the data point and centroid.

3.3.2 Signal to noise ratio (SNR) as a measure of detectability

Assuming that the feature data (summarized in Table 3.7 and Table 3.6) are distributed Gaussian, a measure of *detectability* was presented in [12] for real-time detection of change in cognitive load. Termed as the signal to noise ratio (SNR) in [12], this measure of detectability is a single scale (compared to two scales, i.e. mean and standard deviation) to compare the similarity of two groups of data. When the SNR is high, so is the detectability of data as belonging to one group vs another. The SNR (between two groups of data) is defined as follows.

$$\mathrm{SNR} = \frac{|\mu_1 - \mu_2|}{\max\{\sigma_1, \sigma_2\}} \tag{3.3}$$

$$\mathrm{SNR_{dB}} = 20 \log \left(\frac{|\mu_1 - \mu_2|}{\max\{\sigma_1, \sigma_2\}} \right) \mathrm{dB} \tag{3.4}$$

where

μ_1 = Mean of class 1

μ_2 = Mean of class 2

σ_1 = S.D of class 1

σ_2 = S.D of class 2

3.4 Cognitive load classification results

The data and procedure used for classification are described in this section followed by the analysis using the three classification algorithms discussed in the previous section.

3.4.1 Data fusion

Data fusion approaches aim to improve classification/prediction accuracy using several features. In this chapter we wanted to see if we could improve classification accuracy by fusing two features: PD and RT. For this no specific data fusion strategy was employed, rather the classification algorithms jointly considered these two features for training and testing.

3.4.2 Data structure

Data under analysis consisted of two features i.e. mean response time and mean pupil dilation across 28 participants, as summarized in Table 3.6 and Table 3.7. Features for classification was considered in [RT, PD] pairs (see Table 3.9 for an example).

Table 3.8: **Sample features used by the classifier**

Participant ID	**Mean RT**	**Mean PD**	**Label**
1	697.46	1	CTRL
2	321.56	1	CTRL
3	389.72	1	CTRL
4	306.63	1	CTRL
$\vdots$	$\vdots$	$\vdots$	$\vdots$
27	279.26	1	CTRL
28	275.76	1	CTRL
1	636.23	1.17	2-back
2	539.48	1.27	2-back
3	554.10	1.33	2-back
4	629.52	1.02	2-back
$\vdots$	$\vdots$	$\vdots$	$\vdots$
27	555.83	1.21	2-back
28	304.25	1.46	2-back

The classification analysis is done for one pair at a time, for example, data from CTRL and 2−back as shown in Table 3.8; this binary classification analysis was repeated for all the possible pairs listed in Table 3.9.

Table 3.9: **Classification pairs**

Class1	**Class2**
CTRL	0-back
CTRL	1-back
CTRL	2-back
0-back	1-back
0-back	2-back
1-back	2-back

3.4.3 Procedure for classification:

Following steps were followed for classifying the input features:

1. *Training phase:* In this step we train the classifier on the data that has preassigned class labels. Classifiers were fed with different combinations of training data as shown in Table 3.10; once the data for training is selected, the remaining is used for testing. For example, considering that there were 28 participants, using 80% training and 20% testing meant that data from 22 participant was used for training and the remaining data (6 participants) was used for testing.

2. *Testing phase:* In this step, we use data with unknown labels and try predicting the class or category of the data using the trained classifier.

The features for different combinations of testing and training data were selected randomly. Each combination of testing and training data was repeated 100 times. Accuracy was calculated for each repetition. The overall accuracy of each combination was the mean calculated over the repetitions. Further, it must be emphasized that even though the training data was selected randomly, the same training-testing data pairs were used for the analysis by all three classifiers (SVM, logistic regression and k-means) and the results summarized and discussed in the next three subsections.

Next, we present the results of classification analysis for each type of classifier.

Table 3.10: **Percentage of training and testing data**

Trial Number	Training data(%)	Testing data(%)
1	80	20
2	75	25
3	70	30
4	65	35
5	60	40
6	55	45
7	50	50

3.4.4 Support vector machine

The classification was done for all the possible pairs and the results for each pair is shown as a row in Table 3.11. The first two columns of the Table 3.11 represent the pair of groups under classification and columns 3 to 9 represent classification accuracy for different combinations of testing and training data. According to the average classification accuracy by the SVM in Table 3.11, we can observe two trends: (i) the accuracy did not significantly drop as the training data was reduced to 50 %; and (ii) the accuracy increased, albeit by a small amount, with the difficulty gap in most cases.

Table 3.11: **Accuracy of SVM using using data fusion**

Class-1	Class-2	80-20	75-25	70-30	65-35	60-40	55-45	50-50
CTRL	0-back	85.45	83.57	84.4	84.45	84.72	84.64	85.5
CTRL	1-back	88.54	88.147	87.47	88.35	87.8	88.64	88.85
CTRL	2-back	90.18	89.85	89.05	88.15	88.13	89.04	88.25
0-back	1-back	85	84.92	84.17	83.65	83.81	83.88	83.71
0-back	2-back	89	89.14	87.8	87.5	87.18	86.16	86.96
1-back	2-back	88.63	88.14	88.82	87.2	87.31	87	87.5

3.4.5 Logistic regression

The classification accuracy of logistic regression for each pair is shown as a row in Table 3.12. The first two columns of the Table 3.12 represent the pair of groups under classification and columns 3 to 9 represent classification accuracy for different combinations of testing and training data. One immediate observation of the results in Table 3.12 is that the classification accuracy is much less compared to the SVM classification accuracy that is summarized in Table 3.11.

Table 3.12: **Accuracy of logistic regression using data fusion**

Class-1	**Class-2**	**80-20**	**75-25**	**70-30**	**65-35**	**60-40**	**55-45**	**50-50**
CTRL	0-back	67.2	67.42	67	66	69.7	70.5	67.5
CTRL	1-back	84.18	86.5	85.7	86.8	86.1	86.9	86.5
CTRL	2-back	85.9	87.85	88.47	88.5	88.77	89.04	88.95
0-back	1-back	65.18	66	64.7	64.25	65.36	66.28	66.9
0-back	2-back	73.3	74.57	71.6	74.3	74.5	73.4	74.3
1-back	2-back	65.9	62.3	64.2	63	64.8	65.3	63.5

3.4.6 K-means

Table 3.13 shows k-means accuracy of classification. The first two columns of the Table 3.13 represent the pair of groups under classification and columns 3 to 9 represent classification accuracy for different combinations of testing and training data. Once again, the observation of the results in Table 3.13 is that the classification accuracy is much less compared to both the SVM classification accuracy as well as that by the logistic regression approach, summarized in Tables 3.11 and Table 3.12, respectively.

Next, let us analyze the numbers reported in Tables 3.11, 3.12 and 3.13. Figure 3.18 shows a comparison of binary classification accuracy of all three classifiers considered in this section. Here, the accuracy is averaged for all six classification pairs, i..e, each bar for SVM shows the average of the corresponding column in Table 3.11. From this figure, it clearly shows that the SVM outperforms all the other classifiers

Table 3.13: **Accuracy of k-means using data fusion**

Class-1	Class-2	80-20	75-25	70-30	65-35	60-40	55-45	50-50
CTRL	0-back	45.45	57.1	47	45	50	44	53.57
CTRL	1-back	27.7	57.1	47.05	55	45.4	76	71.42
CTRL	2-back	54.5	35.7	41.1	20	45.4	64	39.2
0-back	1-back	36.3	50	35.2	65	63.6	44	53.5
0-back	2-back	54.5	50	52.9	55	63.6	60	57.14
1-back	2-back	54.5	42.8	52.94	65	54.5	48	42.8

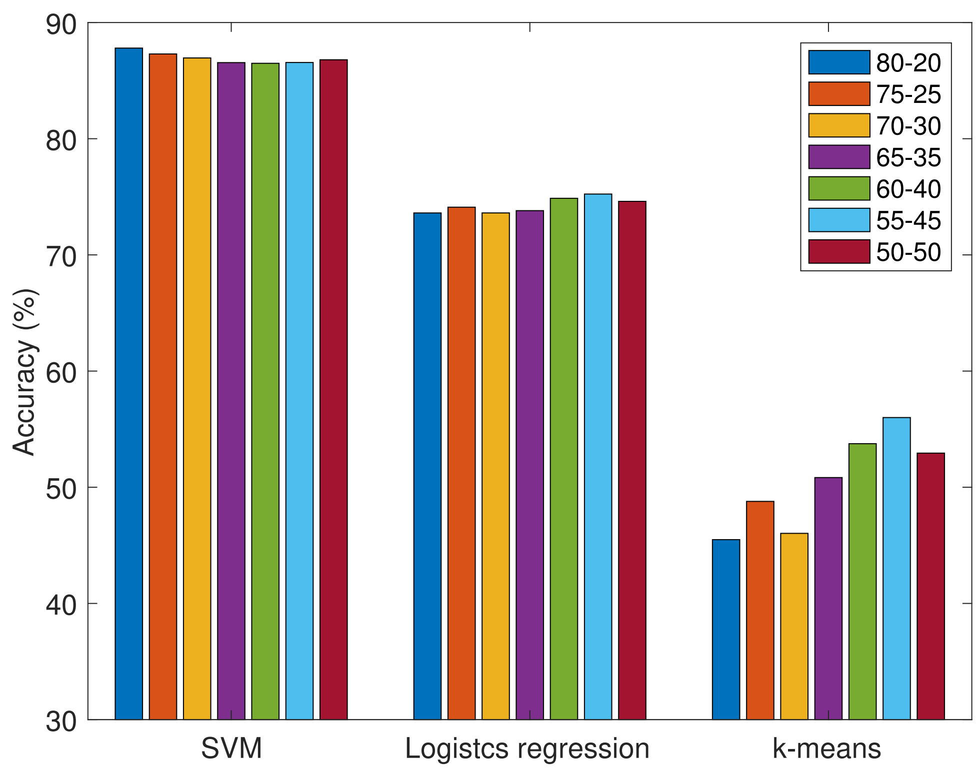


Figure 3.18: **Comparison of classifiers**. SVM classifier has the highest accuracy in comparison to the accuracy of logistic regression and k-means.

in terms of accuracy. Also the classification accuracy did not suffer significantly when the amount of training data was reduced up to 50% (14 participants). Later in section 3.5, we will see that when the training data is reduced below 50% the classification accuracy starts to decrease.

3.5 Performance analysis of individual and fused classifiers

In section 3.4, we discussed the classification performance when features are considered jointly i.e. always PD and RT are jointly considered. Also, since the SVM classifier significantly outperformed the other two classifiers, we will limit our analysis to SVM classifier only. Let us compare the classification performance as follows for further insights:

- Classification using RT only: summarized in Table 3.14
- Classification using PD only: summarized in Table 3.15
- Classification using the fused {PD, RT} pairs (data fusion): summarized in Table 3.11

The objective of such analysis is to understand each feature (PD and RT) in terms of their individual ability to classify cognitive load. The summary of Table 3.14, Table 3.15 and Table 3.11 are illustrated in Figure 3.19. It shows that accuracy remains stable until the training data is reduced to 50% . Beyond that, the accuracy reduces when the training data is reduced. Such behavior is seen to be the same when single features were used as well as when both features were jointly used to classify. Further the PD as a single feature yields the highest accuracy compared to RT alone and PD, RT pairs. The fact that the RT, PD pair yielded less accurately compared to PD alone indicates that, compared to PD, RT did not have any additional information about the cognitive load of the participants. This is an important conclusion that needs to be further studied for a better understanding. Figure 3.20 shows the SNR for pupil dilation and reaction time. Pupil dilation SNR is higher than reaction time SNR. Higher SNR indicates better detectability and hence better classification. Comparison of SNR between the features PD and RT indicate that PD is a better classifier of cognitive load compared to RT. This was also confirmed through classification experiments summarized in Table 3.14 and Table 3.15.

Table 3.14: **Accuracy of SVM using reaction time**

Class-1	**Class-2**	**80-20**	**75-25**	**70-30**	**65-35**	**60-40**	**55-45**	**50-50**	**SNR RT**
CTRL	0-back	85	84	84.3	83.7	84.6	85.2	85	-12.5
CTRL	1-back	87.45	87.3	88.41	88.2	87.4	87.8	88.28	-4.9
CTRL	2-back	89	90.5	88.5	88.9	89.95	88.8	88.4	-3.5
0-back	1-back	83.9	84.5	84.9	85.3	84.05	84.7	83.4	-9.1
0-back	2-back	89.45	87.64	88	87.45	87.3	85.96	87.2	-7
1-back	2-back	89	88.85	88.1	88.35	88	88	87.7	-21.18

Table 3.15: **Accuracy of SVM using PD**

Class-1	**Class-2**	**80-20**	**75-25**	**70-30**	**65-35**	**60-40**	**55-45**	**50-50**	**SNR PD**
CTRL	0-back	87.9	88.7	86.5	86.6	87.4	86.8	86.5	-2.1
CTRL	1-back	92.3	92.7	92	91.5	92	90.3	91.3	4.9
CTRL	2-back	89.27	89.7	89.5	88.7	89.18	89.6	88.6	7.3
0-back	1-back	84.6	85.3	84.5	85.1	84.9	85.3	85.5	-5.3
0-back	2-back	86	86.5	85.2	86.6	85.9	85.84	85.7	-0.2
1-back	2-back	88.7	89.42	89.5	87.9	88.1	88.3	87.8	-5.5

3.6 Conclusions and Discussions

In this chapter, we investigated the classification algorithms for binary classification of cognitive load using multiple features. After comparing the accuracy of all three classifiers i.e. SVM, logistic regression and k-means, SVM was found to have the highest classification accuracy.

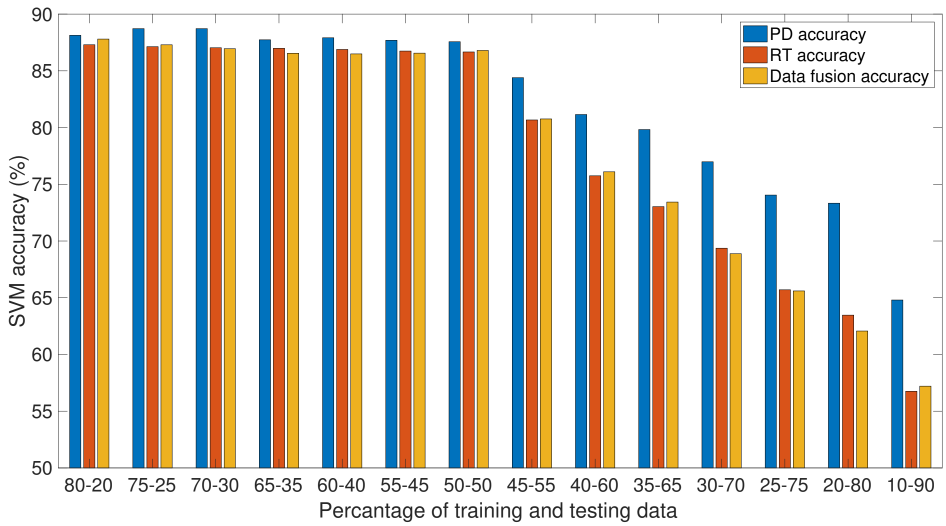


Figure 3.19: **Comparison of classification accuracies**. SVM using only PD and only RT has high accuray as compared to SVM using data fusion. Accuracy is high for all the SVM's when the percentage of training data is high

It was possible to train on 50% of the data to attain the highest classification accuracy with all the three classifiers. Further decrease in the training data resulted in reduced accuracy of the logistic regression and k-means classifier (see Figure 3.18).

For this research classifier with data, fusion was in-effective in terms of attaining high classification accuracy as compared to the accuracy of classifiers with only pupil dilation; reaction time. Further research with more features is required to verify this finding. When the accuracy of SVM using only PD was compared with accuracy of SVM using only RT and data fusion, SVM with PD outperformed the SVM with RT and data fusion. Lower accuracy of SVM using RT suggests that reaction time does not reveal much information about cognitive load. This might be true because the participant performed the DRT task three times in a row which might cause a learning effect. The learning effect makes the task feel less difficulty due to repeated performance. The second reason that might have caused reaction time to be not as significant as pupil dilation is because reaction time was recorded through the button press and pressing a button requires muscle activity. Performing the same

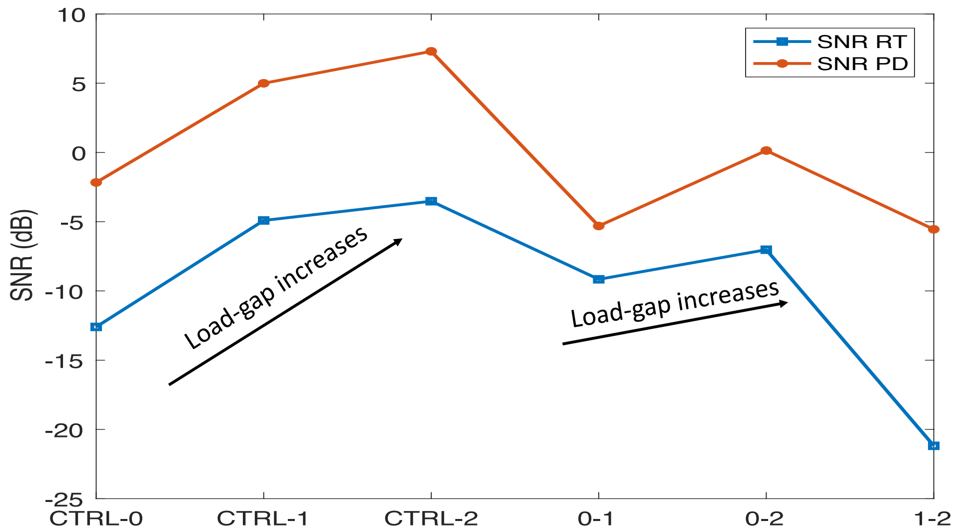


Figure 3.20: **Comparison of SNR, a measure of detectability.** The features PD and RT were compared in terms of SNR. It can be noticed that the SNR increases as the *load-gap* increases i.e. SNR increase along CTRL-0, CTRL-1 and CTRL-2 and it increases along 0-1, 0-2 as well

task creates storage in the muscle memory improving the performance of the task. This finding requires further research which is beyond the scope of this chapter. Infact the comparison also revealed that reaction time might have caused a decrease in the efficiency of the classifier with data fusion, further research is required to have a concrete conclusion about this finding.

The recently found approach SNR proved to be a metric of detectability for different pairs of classes. This finding was confirmed by the high accuracy of the SVM classifier. SNR and SVM classifier both kind of worked in validating each other's findings. A high value of SNR indicated that there is a significant difference between two classes and SVM proved it right by showing high classification accuracy for the particular pair of classes. Similarly, the high classification accuracy of SVM for pair of the class corresponding to the high value of SNR, proved that SNR can effectively detect cognitive load among different pairs of classes. Based on the data reported in this chapter, further studies will be needed to investigate the following three aspects:

1. *Applying signal processing to pupil dilation data:* Developing signal processing approach for improving the detectability of cognitive load through pupil dilation measurements. Raw pupil dilation data is masked by noise which reduces the capability of data to precisely predict cognitive load. Signal processing of the data can help detect cognitive load accurately for multiple difficulty levels.

2. *Investigating missed DRT trials:* Missed DRT trials are defined as the number of times participants failed to respond to DRT vibrations. Analysis of missed DRT trials might reveal significant information regarding the cognitive load.

3. *Analysis of individual learning:* The present approach of this chapter detects and predicts cognitive load based on group learning. Generalizing it for the population is not ideal as the pupil may dilate differently depending on various individual aspects; similarly reaction time may vary from individual to individual. Developing a model based on an individual can help in better detection and prediction of cognitive load.

3.7 Bibliography

[1] "Delayed digit recall (n-back) task." [Online]. Available: http://agelab.mit.edu/delayed-digit-recall-n-back-task pages 15, 52

[2] A. M. Andrew, "An introduction to support vector machines and other kernel-based learning methods by nello christianini and john shawe-taylor, cambridge university press, cambridge." *Robotica*, vol. 18, no. 6, pp. 687–689, 2000. pages 56

[3] P. Ayare, B. Balasingam, F. Biondi, and K. Milne, "On the feasibility of cognitive load detection through pupil dilation measurements," *IEEE Transactions on Human-Machine Systems*, (under review, Aug. 2019). pages 52

[4] J. Beatty, "Task-evoked pupillary responses, processing load, and the structure of processing resources." *Psychological bulletin*, vol. 91, no. 2, p. 276, 1982. pages 49

[5] I. O. for Standardization, "Road vehiclestransport information and control systemsdetection–response task (drt) for assessing attentional effects of cognitive load in driving." 2016. pages 5, 10, 14, 25, 49, 52

[6] D. M. J. Garbade, "Understanding k-means clustering in machine learning," Sep 2018. [Online]. Available: https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1 pages 58

[7] T. Hastie, "Tibshirani, r. and friedman, j.(2009): The elements of statistical learning. data mining, inference, and prediction," 2008. pages 56

[8] C.-W. Hsu, C.-C. Chang, C.-J. Lin *et al.*, "A practical guide to support vector classification," 2003. pages 56

[9] J. Hyönä, J. Tommola, and A.-M. Alaja, "Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks," *The Quarterly Journal of Experimental Psychology*, vol. 48, no. 3, pp. 598–612, 1995. pages 2, 49

[10] D. Jayanthi and M. Bommy, "Vision-based real-time driver fatigue detection system for efficient vehicle control," *International Journal of Engineering and Advanced Technology*, vol. 2, no. 1, pp. 238–242, 2012. pages 50

[11] M. Kutila, M. Jokela, T. Mäkinen, J. Viitanen, G. Markkula, and T. Victor, "Driver cognitive distraction detection: Feature estimation and implementation," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 221, no. 9, pp. 1027–1040, 2007. pages 50

[12] P. Mannaru, B. Balasingam, K. Pattipati, C. Sibley, and J. Coyne, "Real-time cognitive context detection and tracking based on pupil dilation," *IEEE Transactions on Human-Machine Systems*, (under review, Aug. 2019). pages 51, 59

[13] Mathworks. K-means clustering. [Online]. Available: https://www.mathworks.com/help/stats/kmeans.html pages 58

[14] Mathworks. Svm. [Online]. Available: https://www.mathworks.com/help/stats/support-vector-machines-for-binary-classification.html pages 56

[15] N. Nourbakhsh, Y. Wang, and F. Chen, "Gsr and blink features for cognitive load classification," in *IFIP Conference on Human-Computer Interaction*. Springer, 2013, pp. 159–166. pages 50

[16] S. Rafiqi, C. Wangwiwattana, J. Kim, E. Fernandez, S. Nair, and E. C. Larson, "Pupilware: towards pervasive cognitive load measurement using commodity devices," in *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, 2015, p. 42. pages 50

[17] T. A. Ranney, G. Baldwin, L. A. Smith, E. N. Mazzae, and R. S. Pierce, "Detection response task (drt) evaluation for driver distraction measurement application," Tech. Rep., 2014. pages 2, 5, 49

[18] H. Spath, *The cluster dissection and analysis theory fortran programs examples*. Prentice-Hall, Inc., 1985. pages 58

[19] J. Sweller and P. Chandler, "Evidence for cognitive load theory," *Cognition and instruction*, vol. 8, no. 4, pp. 351–362, 1991. pages 50

[20] B. Yin, N. Ruiz, F. Chen, and M. A. Khawaja, "Automatic cognitive load detection from speech features," in *Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces*. ACM, 2007, pp. 249–255. pages 50

[21] R. A. Young, L. Hsieh, and S. Seaman, "The tactile detection response task: preliminary validation for measuring the attentional effects of cognitive load," 2013. pages 10, 49

Chapter 4

Thesis Conclusion

In summary, we have investigated and presented in this thesis two factors contributing towards development of cognitive state adaptive ADAS. Firstly we demonstrated how cognitive load can be detected using psychological measures (chapter 2), validated the findings of pupil dilation and also compared the effect of multi tasking vs single task on cognitive load through pupil dilation. Secondly we presented comparison of cognitive load classifiers along with the effects of using data fusion vs single feature(pupil dilation and reaction time) on the classifier accuracy (chapter 3). A new approach i.e SNR for cognitive load dectability was also demonstrated (chapter 3).

One of the crucial steps in developing a cognitive load measurements is to collect data that helped predicting cognitive load. Analysis of the collected data i.e reaction time and pupil dilation in (chapter 2) showed that:

1. *DRT reaction time can measure cognitive load among multiple levels:* Reaction time is considered as a standard for measuring cognitive load. The collected reaction time during three difficulty levels and CTRL stage (dual experiment) was able to determine cognitive load among two out of six pairs difficulty levels. Reaction time increased with cognitive load (chapter 2).

2. *Pupil dilation collected through non-invasive approach was able to measure cognitive load among multiple difficulty levels:* Validating whether pupil dilation

is capable of measuring cognitive load was one of the important objective of this thesis. Pupil dilation collected during three difficulty levels and CTRL stage(dual experiment) was able to determine cognitive load among three out of six pairs difficulty levels (chapter 2). This finding was validated through DRT, NASA TLX and n-back accuracy.

3. *DRT imposes extra cognitive load other than the primary task:* One of the unique findings of this thesis is that DRT which is a measure of cognitive load imposes extra load on the individual. This finding was demonstrated through comparing the pupil dilation from dual experiment with the pupil dilation from single experiment. Pupil dilation was higher during the dual experiment where the DRT was present as compared to pupil dilation during the single experiment where the DRT was absent (chapter 2).

After successfully detecting the cognitive load the next goal was to build a model that can predict cognitive load accurately. The analysis of the classifiers showed that:

1. *Support vector machine has the highest accuracy among the compared classifiers:* After analyzing there different binary classification algorithms for cognitive load support vector machine has the highest accuracy (90%) as compared to logistic regression and k-means.

2. *Pupil dilation has more information regarding cognitive load as compared to reaction time:* Accuracy of the classifier with pupil dilation was high compared to the accuracy of classifier with reaction time and data fusion. Pupil dilation outperforms reaction time and data fusion technique for classifying cognitive load.

The culmination of the works presented in this thesis serves as a potential benchmark in a niche technological field that is presently under-developed. Making an driver assistance system that accommodates it's operation depending on his/her cognitive state is the need of the hour. We imagines a future in which the ADAS system could be modulated automatically instead of keeping it in the hands of the vehicle driver. Existence of such a system will not only help keep the driver safe but also will ensure the safety of the fellow road companions. With the help of application-specific systems such as the works presented in this thesis, a more widespread accessibility of cognitive load detection and classification will certainly prove to benefit the general public, industry, and academia alike.

In terms of future work, we suggests replicating the same experiment in real-time driving environment. Presently the findings correspond to a laboratory environment. Real time driving environment can introduce a variety of new variables which need to be accommodated. It has been established that pupil dilation (from non invasive techniques) can effectively measure cognitive load. This provides an opportunity to use pupil dilation instead of DRT which is a invasive technique for measuring cognitive load in driving environment for future work. The cognitive load classification in the present work is based on group learning it will be more realistic in terms of applicability to consider individual learning effect. Exploring the other features of DRT such as the missed trials might help improving the accuracy of classification for cognitive load. Lastly developing a model for real time cognitive load detection based on psychological measures.

Vita Auctoris

NAME:	Prathamesh Ayare
PLACE OF BIRTH:	Mumbai, India
YEAR OF BIRTH:	1993
EDUCATION:	SBMP, Mumbai , India 2009-2013, Diploma Electrical Engineering
	Atharva College of Engineering, Mumbai, India 2013-2016, Bachelor of Engineering Electrical Engineering
	University of Windsor, Windsor, Ontario 2017-2019, Master of Applied Science Electrical Engineering