THINK2ACT: USING MULTIMODAL DATA TO ASSESS HUMAN COGNITIVE AND PHYSICAL PERFORMANCE

A Dissertation

Presented to the Department of Computer Science and Engineering

at the University of Texas at Arlington

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Maher Abujelala

December 2019

THINK2ACT: USING MULTIMODAL DATA TO ASSESS HUMAN COGNITIVE AND PHYSICAL PERFORMANCE

Maher Abujelala

University of Texas at Arlington 2019

ABSTRACT

As computers become more advanced, affordable, and smaller in size, we start to use them in almost every aspect of our daily life. Nowadays, the use of computers is not just limited to accomplish work-related tasks. Instead, we use computers for education, entertainment, healthcare, and in many other areas to facilitate our daily life activities. From here, the Human-Computer Interaction (HCI) field emerged. HCI is a multidisciplinary field of study that focuses on utilizing computers and technology to interact with humans, improve their quality of life, and enhance their performance. The rapid advancements in other related research fields, such as robotics, artificial intelligence, and sensor technologies, have tremendously improved human-computer interaction applications and made it more personalized, adaptive, and smarter.

This research explores innovative HCI applications using robotics, sensors, and wearable technologies to monitor and assess human cognitive and physical abilities. The systems collect, analyze, and evaluate multimodal data, which include system-specific metrics and data from non-invasive sensors. The sensors used in this research include electromyography (EMG), electroencephalogram (EEG), electrocardiogram (ECG), electrodermal activity (EDA), pulse oximeter, inertial measurement unit (IMU), eye-trackers and cameras. The analysis and evaluation of these multimodal data are done using statistical analysis and ma-

chine learning techniques.

Primarily, this dissertation discusses gamified robot-assisted assessment and rehabilitation, the impact of sleep quality on cognitive performance, the importance of understanding human behavior and physiology to provide adaptive and personalized training, and the impact of cognitive workload on human physical performance. The outcome of this research is a multimodal cognitive and physical assessment platform, called ”9PM” that stands for 9-Peg Moves. The platform combines a simple physical task based on the principles of a standard upper extremity test, with other standard cognitive tests to assess user cognitive and physical performance and understand the correlation between users' performance and their physiological and behavioral responses.

BIOGRAPHICAL SKETCH

Maher Abujelala was born in Tripoli, Libya, in 1992. In 2009, he finished his engineering high-school education at Benghazi Center for Gifted students (previously known as Al-Fateh Center for Gifted students) in Benghazi, Libya, and he achieved the 4th highest grade in the national examination across the country. Following that, he received a 4-year full scholarship from the Libyan Ministry of Education to study undergraduate degree in the USA. In 2010, he joined Washington State University (WSU) at Pullman, Washington, to do his undergraduate degree in Computer Engineering, which he finished in May 2014.

Maher started his Ph.D. degree in Computer Engineering at the University of Texas at Arlington (UTA) in January of 2015, and he defended his Ph.D. thesis in November 2019. He worked as a researcher at Heracleia Lab, a human-center computing laboratory, under the supervision of Prof. Fillia Makedon. During his time at Heracleia lab, he had the opportunity to work as a graduate research assistant on several NSF-funded projects.

Maher's research interests revolve around Human-Computer Interaction, focusing on physiological and wearable sensors, and healthcare applications. Maher has accepted an offer to be a Postdoctoral Associate at the Department of Psychiatry at Yale University, under the supervision of Prof. Morris Bell. At Yale University, Maher will work on an NSF-funded project for assessing executive function in children, and he will be involved in the development of machine learning and deep learning methods for this project.

To my beloved parents and kind-hearted sisters for their unconditional love, support, encouragement and patience.

ACKNOWLEDGEMENTS

First of all, I would like to express my sincere gratitude to my supervisor Prof. Fillia Makedon for her continuous support and encouragement during the last five years. Also, I would like to thank my colleagues at Heracleia lab for making my stay at UTA very productive and joyful. Special thanks go to Akilesh Rajavenkatanarayanan, Varun Kanal, and Konstantinos Tsiakas. I am very fortunate for having a caring supervisor and surrounded by smart, hospitable colleagues. Also, I would like to thank my supervising committee, Dr. Eric Becker, Prof. Junzhou Huang, and Prof. Chengkai Li, for their valuable comments and feedback. Last but not least, I would love to thank my parents, Muftah and Mariam, and my three sisters, Nabila, Hekma, and Amira, for their endless love, support, encouragement, and patience.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

HUMAN ACTIVITY ASSESSMENT AND TRAINING: TRENDS AND CHALLENGES

1.1 Introduction to Human Abilities Assessment

To solve a problem effectively, we need to find out what causes it in the first place. Every year thousands of workers have unforeseen accidents and injuries in the workplace. Is that due to the lack of safety measures, lack of proper training, or that the workers are either physically or cognitively incapable of performing the job? To answer such a complex question, we need to monitor the work environment, and understand if it is a human error. The workers performance needs to be assessed, and based on that, they should receive proper training to improve their weaknesses. The abilities of the workers vary, so they need to be assessed and given personalized training. Unfortunately, it is time-consuming and challenging for human experts to provide such assessment and training for every worker. This issue is not just related to workers, but it also touches students, patients, soldiers, and many others. To overcome this issue, researchers come up with standardized tests to help assess users and determine their weaknesses. Generally, these tests are not to make formal diagnoses, but to identify the deficiency and difficulties the participants have. One of the well-known initiatives to develop such tests is the National Institutes of Health Toolbox (NIH Toolbox) initiative, which is part of the NIH Blueprint for neuroscience research initiative [86, 37]. The NIH Toolbox includes over 100 stand-alone measures to assess cognition, emotion, motor, and sensation for children (3 years old and above) and adults. Also, it is designed to be brief, affordable, and usable in large-scale studies.

The next sections discuss the current trends in cognitive and physical assessment and the motivation and structure of this dissertation.

1.2 Cognitive Assessment

Cognitive assessment is one of the initial procedures to discover mental impairments and learning disabilities. In this context, cognitive assessment refers to the evaluation of mental capacity to gain and illustrate knowledge and comprehension, which includes the ability to think, do problem-solving, memorize, execute complex behaviors, and many others. One of the commonly used cognitive assessment tests in cognitive psychology is the Stroop task, which is considered to be the golden standard in attentional measures [71]. The participant in this assessment sees color words, and the font-color of the word is different. The participant is supposed to name the font-color instead of reading the word. For instance, if the color word is Red, the participant should say 'Blue' rather than 'Red.' This task shows the participant's mind control ability. Reading the words is considered an automatic process, but naming colors is regarded as a controlled process that requires more attention. In [45], the Stroop task was used to identify executive function deficits associated with attention deficit hyperactivity disorder (ADHD) in children and adolescents. Their results show that poor performance in the task is not sufficient to diagnose ADHD, but they found that children and adolescents with ADHD consistently exhibit poor performance. This finding confirms that people with ADHD generally have difficulties with attention and concentration. Some researchers believe that the Stroop task is sensitive to brain damage, especially when it is in the frontal or prefrontal area [45]. [47] found using Magnetic Resonance Imaging (MRI) that children with ADHD had significantly smaller right frontal width than control subjects.

In contrast, [21] explored the inhibitory control in children with Autism Spectrum Disorder (ASD) using three cognitive tasks (Stroop task, Flanker task, and Go/No-go task). There is a fixed number of stimuli shown at the same time (e.g., OOXOO) in the Flanker task. The participant should identify if the stimuli 'flanking' the target stimulus (e.g., the middle stimulus) have the same, opposite, or a neutral definition. In contrast, the Go/No-go task has one stimulus shown at a time. The participant should act (e.g., press a button) when a specific target stimulus (e.g., green light) is presented but should stay still otherwise. They found that children with ASD had a similar performance to the control subjects in the Stroop task, but they performed more poorly than the controls in the Flanker and the Go/No-go tasks.

The cognitive assessments discussed above focus mainly on cognitive impairments that affect young participants. However, there are other cognitive dysfunctions commonly affect the elderly, and researches try to make assessments for them. For instance, the Clock Drawing Test and the Mini-mental State Examination are dementia screening assessments for the elderly.

1.3 Physical Assessment

Physical assessment in this context primarily refers to the evaluation of motor function, dexterity, strength, power, and endurance. Physical assessment of post-stroke, traumatic brain injury, or multiple sclerosis patients helps clinicians understand the level of physical dysfunction and create a treatment plan for faster recovery or to slow down the impairment progression. Some of the commonly used physical assessment tasks are the 9-Hole Peg test, Box and Blocks test, Timed Up-and-Go test, and the Manual Muscle Testing [99].

The 9-Hole Peg and the Box and Blocks are manual dexterity tests that assess upper extremity function. The 9-Hole Peg is considered a gold-standard test. In this test, the patient is instructed to move 9-pegs from a container to a block with nine holes, then back to the original container. The patient is timed and needs to do the test twice using each hand. Similarly, in the Box and Blocks test, the patient is asked to move blocks quickly from one side of a rectangular box to the other side in 60-seconds. The test score is the count of the blocks the patient moves. The Timed up-and-Go test is a mobility assessment to assess balance and walking defects. The patient in this assessment is instructed to sit back in a standard armchair. Once the assessment starts, the patient needs to stand up, walk a distance of 3 meters, and come back. The Manual Muscle Testing is one of the methods of measuring extremity muscle strength. In this test, the patient is asked to hold the corresponding body part while the practitioner exerts manual resisting force in the opposite direction, and the muscle strength is then evaluated accordingly.

1.4 Motivation

The advances of sensor technology have led to new cost and power-efficient devices in the market, making them available for a larger population of developers and researchers. Nowadays, many portable sensing devices are consumer-grade and targeted for lifestyle applications. Smartwatches and smartphones are just two of the many tools that can embody sensors that can track our movement, sleep pattern and quality, heart-rate data, and many other possibilities. Leveraging this vast amount of data can help us understand more about our bodies and capabilities. It can be used to advance healthcare and educational applications by personalizing them based on our needs. However, assessing the skills of an individual is not straightforward. Human experts can have the ability to evaluate users intuitively by identifying the skills required for evaluation; however, developing systems and algorithms to do that is a complicated task. With that in mind, researchers try to incorporate sensors and technology in the assessment to make it quantifiable and easy to administrate. For instance, they can use technology and software to automatically time experiments and collect score and user feedback instead of using a stopwatch and write notes by hand. In recent years, there have been many efforts to use sensors to detect human mental and physical state. For instance, researchers have worked on using EMG signals to detect muscle fatigue [110], EEG signals to sense pain, engagement, and anxiety [56, 19], and ECG and EDA data to detect mental stress [103, 106].

Our work moves towards the design of an intelligent system that acts autonomously to decrease the workload on experts and allow them to monitor the user's performance and plan on the next course of action accordingly. For this reason, this dissertation investigates how physiological and behavioral data can be used to assess the user's cognitive and physical performance.

1.5 Dissertation Structure

This section shows the structure of this dissertation and the main focus of each chapter:

- Chapter 1 introduce the trends and challenges in human assessment, and shows some of the commonly used assessment.
- In Chapter 2, we discuss robot-assisted assessment and rehabilitation related work. Moreover, we showcase our work in analyzing EMG data to detect muscle fatigue during physical rehabilitation [54]. In addition, we show one of our systems that combine cognitive and physical assessment using serious games and a rehabilitation robot [20]. This chapter presents parts of our previous publications [54] and [20].
- Chapter 3 focuses on sleep-related studies and the impact of sleep quality and disorder on our performance. We showcase our pre-screening tool for apnea in a home environment that utilizes wearable sensors and computer vision techniques. This chapter presents parts of our previous publication [55].
- In Chapter 4, we discuss brain-computer interfaces and show our system that uses EEG data to monitor user's engaged enjoyment. This chapter presents parts of our previous publication [1].
- In Chapter 5, we discuss our novel multimodal cognitive and physical assessment platform, called "9PM", and how we utilized multimodal data analysis in assessment.
- Finally, in Chapter 6 we show the concluding remarks and future directions of this dissertation.

CHAPTER 2

ROBOT-ASSISTED ASSESSMENT AND REHABILITATION

2.1 Introduction

Shoulder dislocation, stroke, or traumatic brain injury (TBI) are a few of the conditions that can be the result of an accident, or physiological or neurological dysfunction. These cases can limit people from performing daily life activities like tying shoes and going for a walk. Around 20 million people worldwide suffer from strokes each year, and 795,000 of these cases happen in United States [18, 78]. That costs the United States approximately 34 billion a year [38]. Five million out of the 20 million people die because of stroke, and only 15 million people survive. Third of the people who survive live with permanent after-effect disabilities. That makes stroke one of the significant causes of death and disability. Luckily, not all of these disabilities are permanent, and some of them can be mitigated with physical rehabilitation exercises.

Conventional rehabilitation has proven to be very useful. It requires the patients to visit a rehabilitation center to see a therapist. However, that makes rehabilitation labor-intensive and inconvenient to patients who have to travel for long-distances. Many traditional rehabilitation exercises require the patients to perform repetitive, simple activities such as stretching their arm and moving it in circles. The repetitive nature of these exercises might make them less engaging. Some of the research studies in the last decade have focused on resolving the issues related to conventional rehabilitation. For example, these studies explore robotic and gamified rehabilitation to reduce the workload on the therapists, allow for telerehabilitation, and make the rehabilitation procedure more fun and engaging [48, 112]. Gamified-rehabilitation and robotic-rehabilitation generally utilize robots (e.g., robotic-arm, exoskeleton, etc.) to help the patients perform exercises in a game-based, virtual-reality environment. This engagement is essential to improve the patient's condition as studies have shown that patients who get engaged in the rehabilitation exercises have a better chance of retaining the white matter integrity in their brain [13].

2.2 Background

This section provides brief background information about robot-assisted rehabilitation, gamified rehabilitation, and fatigue.

2.2.1 Robot-Assisted Rehabilitation

Robots are generally used to do tedious, repetitive tasks. That makes them suitable for rehabilitation exercises. As mentioned earlier, many of the rehabilitation exercises require the patients to repeat them for long periods to improve motor function. Having a robot to help the patient repeat these exercises could reduce the rehabilitation cost and the workload on the medical staff, and allow the patient to do the exercises remotely at home. Some robots nowadays have embedded sensors that can record the patient's performance. For instance, the robot can accurately record the patient's movement speed, and the amount of the force the patient applies. It is very beneficial to record these metrics, observe the patient's progress over time, and evaluate the effectiveness of the rehabilitation plan.

Various types of robots are used in healthcare applications. Surgical robots, socially assistive robots, and rehabilitation robots are some of these robots. For instance, surgical robots are used to perform remote surgeries and surgeries requiring high precision (e.g., retinal detachment surgery). Socially assistive robots interact with patients in a friendly way, and researchers usually utilize them for cognitive assessment and training tasks [113]. On the other side, rehabilitation robots help patients gain their motor function and perform physical exercises. For instance, multiple sclerosis (MS) patients in one of the studies were asked to use an exoskeleton robot and a treadmill for gait rehabilitation [10]. The study found this method significantly effective, and it could help patients with severe walking disabilities to improve their walking distance, velocity, and knee-extensor strength. Another study compared the usefulness of a 6-degrees of freedom (DOF) robot manipulator to conventional therapy techniques in upper-limb physical rehabilitation [70]. The researchers found that their post-stroke patients who performed robot-assisted rehabilitation had significant improvement, in the first two months, in strength, reach, and an overall decreased impairment.

2.2.2 Gamified Rehabilitation

Gamified rehabilitation uses rehabilitation exercises in the form of a game. It makes rehabilitation fun, and it keeps the patients engaged in the rehabilitation activities for more extended periods. [79] enrolled post-stroke patients in a gamified rehabilitation program for two weeks. Their findings suggest that their rehabilitation program was useful for upper limb rehabilitation. Among the games used in rehabilitation are serious games. Serious games are usually computer games designed for purposes other than pure entertainment [101]. Such games have been tremendously useful in various fields, such as education, military training, and healthcare.

2.2.3 Fatigue

Fatigue is a subjective phenomenon that is studied in many fields, such as medicine, neurology, and psychology. Still, it does not have a universally accepted definition [115]. One of the fatigue's popular definitions is by Jensen *et. al.* [51], the definition states:

> *"Acute fatigue is most often caused by excessive physical or mental exertion and can be relieved by rest. Normative fatigue may be influenced by circadian rhythm and results from the activities of daily living, while chronic fatigue is most often prolonged by stress or tension on the body and is less likely to be relieved by rest alone."*

With that in mind, this section focuses on acute muscular fatigue. The frequency of muscle exercise in physical rehabilitation leads to physical fatigue. Physical fatigue could force the patients to alter the rehabilitation exercises to avoid the pain, which could negatively impact the effectiveness of therapy [88]. Fatigue is usually detected using subjective measures (e.g., self-reporting [17, 114]) or objective measures (e.g., physiological responses [82]). For example, [85] used EEG, heart rate variability, and blood biomarkers to detect fatigue while driving. Section 2.4 discusses the use of EMG data to detect fatigue. Mean and median frequencies of the EMG data are found to be effective in detecting physical fatigue [22, 27].

2.3 Multimodal Analysis of Serious Games for Cognitive and Physiological Assessment

This section describes our robot-assisted rehabilitation system [20]. This system utilizes serious games to gamify robot-assisted rehabilitation and make it engaging and cognitively impactful. The users of the system complete challenging tasks using the Barrett WAM arm (see Figure 2.1), and the system records their hand-eye coordination using the Eyetribe eye-tracker (see Figure 2.2).

Figure 2.1: Serious Games: Barrett WAM Arm.

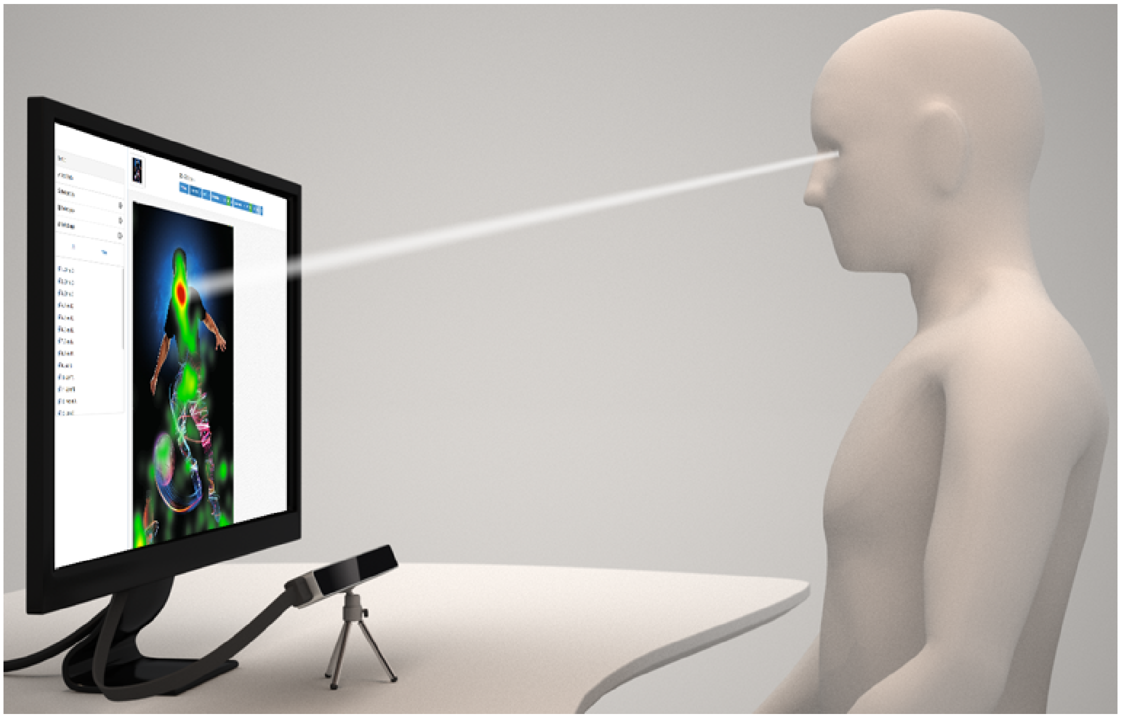


Figure 2.2: Serious Games: Eyetribe Eye-Tracker [34].

2.3.1 Tasks

There are four tasks in this study that the user needs to perform using the robotic arm. In the first task (L1), the user should paint a virtual fence using the robotic arm, see Figure 2.3-A. The fence is divided into ten sections, and each section should be printed differently. There are five different paint buckets the user can use. To figure out which paint color to apply for each section, the user must read the instructions on the screen. The next instruction will not appear on the screen until the user moves the robotic arm back to the center of the screen. The instruction is the name of one of the five available paint colors. The font-color of the instructions in this task is white. Then, the user goes shopping in a virtual supermarket. In the second task (L2), the user is given a list of five items to pick from a single aisle in the supermarket, see Figure 2.3-B. The user has three

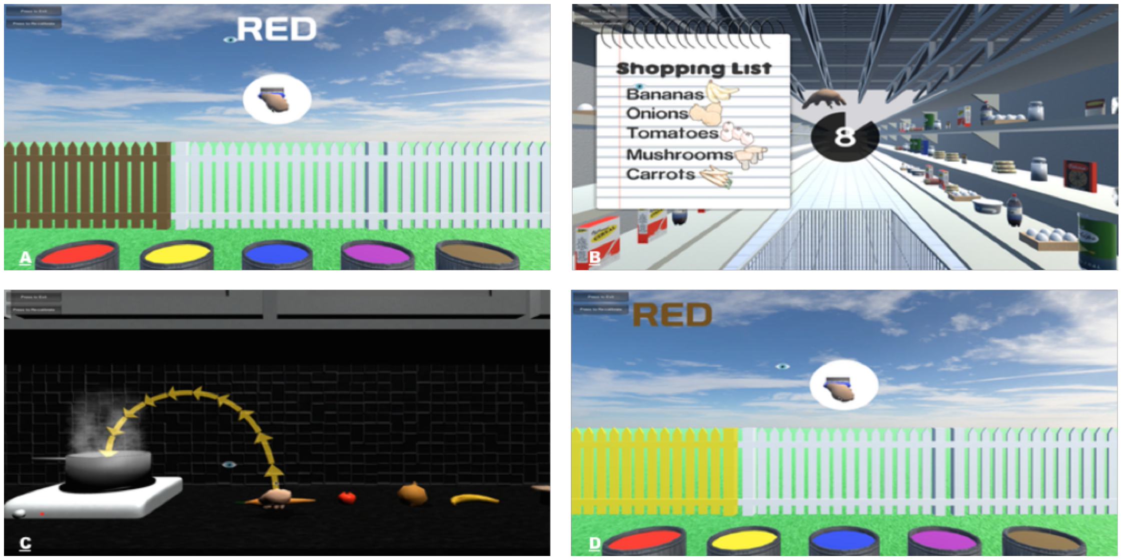


Figure 2.3: Serious Games: (A) Painting a Fence (L1) (B) Supermarket (L2) (C) Cooking a Meal (L3) (D) Painting a Fence (L4) [20].

attempts to collect all the items. In the first attempt, the 5-items appear on the screen for 15 seconds, and the user needs to memorize them. Then, the user goes through the aisle and picks up the items by moving the robotic arm in the 3D environment. The user receives positive points for collecting the required items and negative points for picking up the wrong items. If the user could not pick up the 5-items in the first attempt, then he/she can go for a second and a third attempt. In the second attempt, the remaining items are presented for 5 seconds, and then the user goes through the aisle again. However, this time, the shopping cart goes slower than before. The third attempt is similar to the second attempt. There are visual cues (i.e., flashing items) on the shelves to help the user recognize and pick up the correct items in the third attempt. After that, the user uses the 5-items listed on L2 to cook a virtual meal in the third task (L3), see Figure 2.3-C. The 5-items are available on a kitchen table, and the user needs to pick them up using the robotic arm. After picking an item, the user needs to follow a predetermined 3D trajectory to move the item to the cooking pot. The user repeats this procedure for all the items, and each item has a different 3D trajectory. Then, the user goes back to paint the fence in the fourth task (L4), see Figure 2.3-D. In L4, the font-color of the instructions can be any one of 5-paint colors. The semantic meaning of the instructions and their font-color might be different. The user should paint the fence based on the font-color of the instruction rather than its semantic meaning.

L1 and L4 are two versions of the Stroop test, and they are used to test the user's inhibitory control and assess his/her cognitive ability [108]. The Stroop test is discussed details in Section 5.2.2. L2 is based on a model from Baddeley [7], and it is used to test the user's working memory. We surmise that the number of attempts the users need to collect the 5-items represents their working memory ability. The predetermined 3D trajectories in L3 make the rehabilitation exercises more challenging and can indicate the user's hand-eye coordination ability [65, 93].

Table 2.1: Serious Games: Participants Demographics [20].

Participant Demographics - N = 12 Participants
Age Range: 18-29
Gender: 6 Males, 6 Females
Handedness: 12 Right, 0 Left

2.3.2 User Study and Experimental Procedure

Twelve people, between the age of 18 and 29 years old, participated in the study. Table 2.1 summarizes the demographic information of the participants. The study personnel provided the participants with verbal instructions and clarified any questions they may have. Then, they read and signed the consent form. The first phase of the study was the eye-tracker calibration process. Nine points flashed on the screen, one at a time, and the user needed to follow them with his/her eyes only. The calibration process has four values: *perfect*, *good*, *moderate*, and *poor*. If the calibration value is moderate or poor, the participants need to repeat the calibration process. However, if the calibration does not improve, the participants need to continue the study, and their data would be excluded from the analysis. Only 4 out of the 12 participants were able to have a perfect or good calibration value. We suspect that the deviation in the participants' facial features is the reason for the inconsistency in the calibration process. Then, the participants tried trials of the tasks. The trials are short versions of the tasks. After that, the participants performed the tasks in order (i.e., L1, L2, L3, and then L4). Figure 2.4 illustrates the experiment setup. Also, this online video [1] shows the eye-tracker calibration process, the trials, and the four tasks.

2.3.3 Results and Discussion

There were multiple issues with the eye-tracking device that affected the recorded data. We found that the study environment should be controlled. For example, the change in the room lighting condition affects the pupil dilation

[1] $https://www.youtube.com/watch?v=8TPn_aqfVi8$

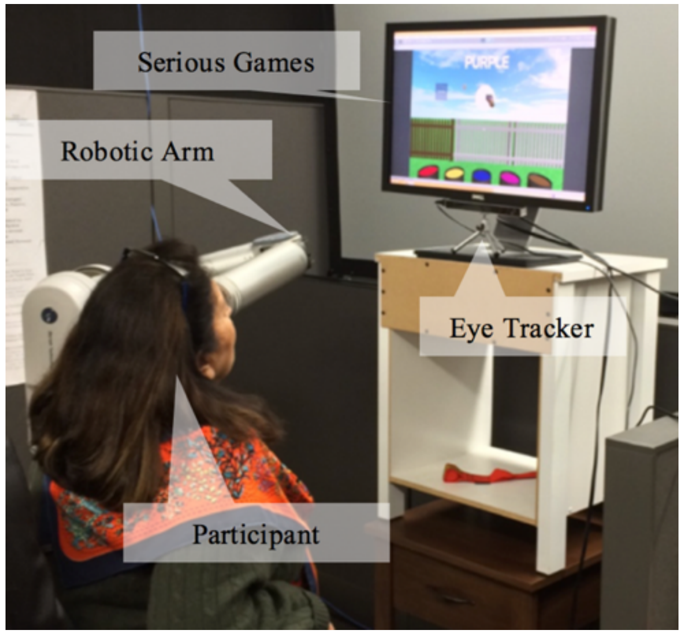


Figure 2.4: Serious Games: Experimental Setup and a participant [20].

of the participants. Therefore, we had to perform the study in a room without windows to make sure that the lighting can be easily controlled. We also found that the participant's sudden movements and far distance from the eye-tracker (over 60cm) affected the quality of the eye-tracker data. To allow the participants to be comfortable while performing the combined cognitive and physical activities, we tried to limit the experiment restrictions and provide the participants with autonomous control. However, that was not favorable for the pupil data acquisition. [75, 76] were able to get more reliable pupil data by limiting the participants' movement.

The results of the valid data from the four participants are listed in Table 2.2. Also, Figure 2.5 shows the eye-delay and move-delay data per participant. The eye-delay is the time the participant takes to recognize a new instruction, and the move-delay is the time the user undertakes to act (i.e., go to the paint bucket). Using a Pearson's r correlation, we found that there is a positive corre-

Table 2.2: Serious Games: Results [20].

Participant	Correctness(L1)	Iterations(L2)	Eye-Hand Coordination (Average)(L3)	Correctness(L4)
1	90%	2	83%	100%
2	100%	2	76%	90%
3	100%	3	78%	100%
4	100%	2	72.5%	90%

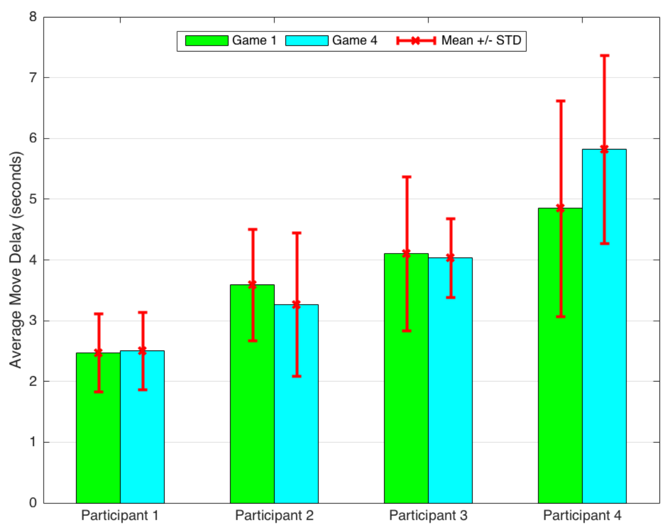


Figure 2.5: Serious Games: Average Move Delay for L1 and L4 [20].

lation between the participant's eye-delay and move-delay, as shown in Figure 2.6. It is a strong, positive correlation with $r(78) = 0.89, p \leq 0.001$. That shows that the faster the participants' eyes reacted to new instructions, the faster their physical response was.

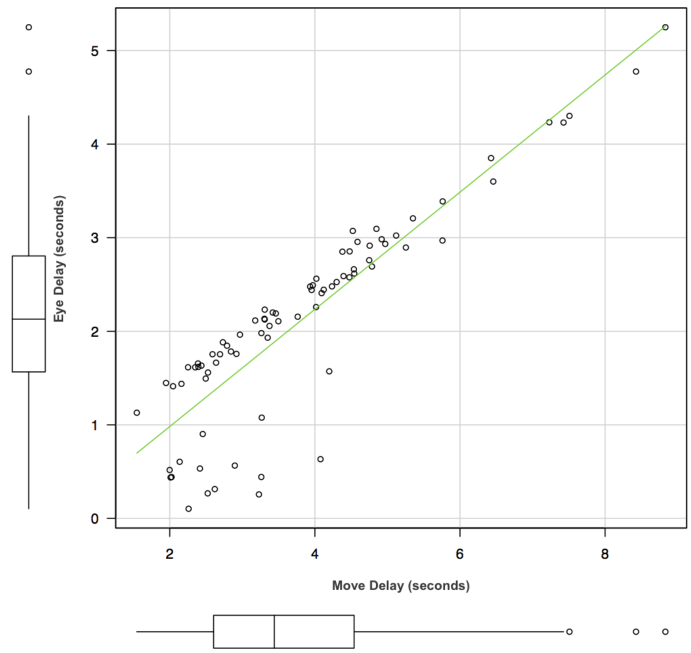


Figure 2.6: Serious Games: Pearson's correlation between Move & Eye Delay [20].

2.4 Adaptive Robotic Rehabilitation using Muscle Fatigue as a trigger

One way to reduce physical fatigue during robotic rehabilitation is by having the robot assist the user to complete the exercise. For instance, [92] uses a robot in human-robot co-manipulation tasks. The robot learns skills from human demonstration. When the robot detects fatigue from EMG data, it completes the task based on these skills to allow the user to rest and recover. Another way to reduce fatigue during physical rehabilitation in real-time is by reducing the forces a robot can exert on the user, which is the topic of this section. In this section we describe our work [54]. In this work we designed a muscular fatigue detection instrument that analyzes EMG data during robot-assisted rehabilitation to detect fatigue. Initially, the robotic arm in the rehabilitation session exerts forces challenging the user. The robot does that by exerting forces in the opposite direction to the arm's movement. When the system recognizes fatigue from the EMG data, the robotic arm starts assisting the user to complete the exercise, instead of challenging him/her.

Figure 2.7: Adaptive Robotic Rehabilitation: Delsys EMG sensor (front and back) [28].

2.4.1 Devices and Exercises

This robotic rehabilitation system has two primary devices: a robotic-arm (Barrett WAM arm, see Figure 2.1) and an EMG sensor (Delsys EMG sensor, see Figure 2.7). The Barrett WAM arm is a 4-DOF robot with a spherical handle that is used as an end-effector. The robot is programmed to restrict the user's movement to a specific trajectory that can be manually recorded before every exercise. The ability to record different paths is advantageous as it allows the system to be used with various exercises and with different users, with disregard to their height and body structure. The robot can also exert forces on the user's hand/arm, similar to the effect of holding weights in rehabilitation exercises. On the other hand, the Delsys EMG sensor is attached to the user's arm using a double-sided adhesive to collect EMG data from the major muscle responsible for the movement.

There are multiple upper-limb rehabilitation exercises. In this study, we considered three exercises that the user can perform while sitting upright on a stool and holding the robot's end-effector. These exercises are Shoulder Flexion (SF), Shoulder Abduction (SA), and Elbow Extension (EE) [52, 40, 33]. To induce physical fatigue in a short time, the users performed the exercises with their non-dominant hand. To do SF, the users raise their arm in front of them, to the overhead level, and hold it straight. For SA, the users raise their arm by their side, to the shoulder level, and keep it in place. During EE, the users extend their arm backward, and then they lean forward. During SF and SA, the EMG data is obtained from the deltoids of the non-dominant hand [53, 67], but it is obtained from triceps during EE [58]. In these exercises, the users move the robot's end-effector from the start-point to the end-point and hold it in place at the end-point. Figure 2.8 shows the three exercises. The start-points are marked in green in the figure while the end-points are marked in red.

2.4.2 Preliminary and Evaluation Studies

In this project, two studies are conducted: a preliminary study and an evaluation study. In the preliminary study, we collected the EMG data needed to make a fatigue detection algorithm. The evaluation study was used to evaluate the performance of the algorithm and the adaptive rehabilitation system.

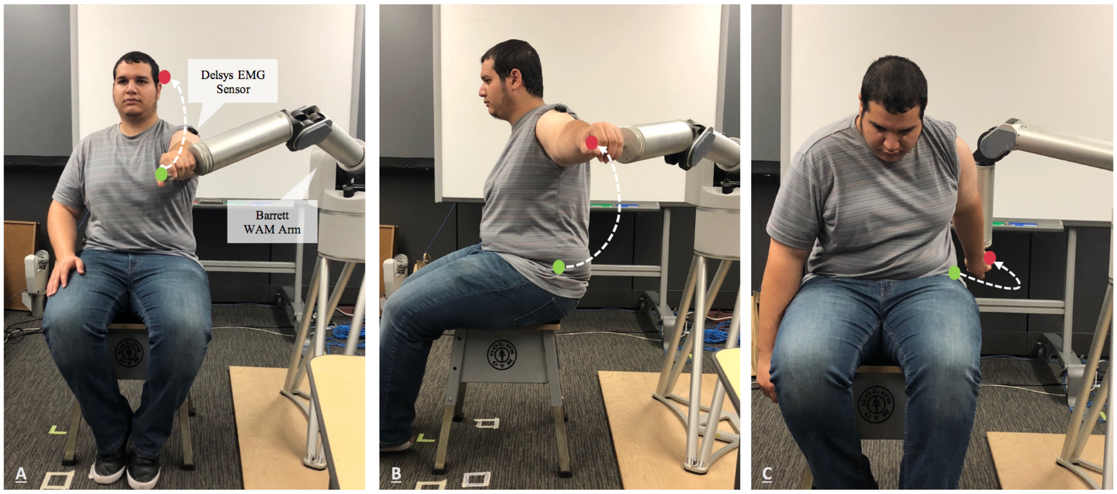


Figure 2.8: Adaptive Robotic Rehabilitation: A- Shoulder Flexion (SF) B- Shoulder Abduction (SA) C- Elbow Extension (EE). The start-points are marked in green while the end-points are marked in red. [54].

The experiment workflow is very similar in both the studies. The participants start by watching a video demo [2] explaining the experiment and illustrating the three exercises. Then, they read and sign the consent form. The study personnel is available to answer any questions the participants may have. This procedure takes a few minutes, and it could help the participants relax and alleviate any physical stress or fatigue they were experiencing before the study. After that, the participants perform the three exercises in order, SF, SA, and EE, respectively. At the beginning of each exercise, participants hold the robot's end-effector and move it from the start-point to the end-point to record the exercise path. The path recording confirms that the end-effector is within the participant's reach and that the participant can perform the exercise smoothly. It also helps the robotic arm apply resistive forces to the participant arm movement. The participant then takes the end-effector slowly back to start-point. When the

[2] $https://youtu.be/bYiix8KGqYs$

participant reaches the start-point, the study personnel adds resistive forces to the robotic arm equal to 20 newton. The resistive forces try to pull the participant's hand down through the recorded path to the start-point. The participants then are asked to try to move the end-effector to the end-point to check if they feel comfortable. Based on that, the study personnel adjusts the resistive forces to make sure that it is safe for the participant to perform the exercise. Then, the study personnel helps the participants place the EMG sensor on their arm. While the resistive forces are applied to the robotic arm, and the EMG data is recorded from muscles, the participants do the exercises. They do the exercises by moving the end-effector from the start-point to the end-point. They hold the end-effector at the end-point, and they are instructed to immediately notify the study personnel when they start feeling fatigued and not to wait until they cannot hold the end-effector anymore. After every exercise, the participants relax to relieve the fatigue and avoid its cascading effect.

Preliminary Study

Ten healthy participants participated in the preliminary study, and they performed the three exercises. They held the end-effector at the end-point until they felt fatigued. The study personnel marks that time the participants report the fatigue, stops recording the EMG data and asks them to move the end-effector back to the start-point slowly. The marked time is used as the ground truth of subjective fatigue, and in conjunction with the EMG data, it is used to design the muscular fatigue detection algorithm. The details of the algorithm are available in [54]. The participants repeated the same procedure for all three exercises.

Evaluation Study

The evaluation study included 20 participants (15 male, 5 female), with a mean age of 25.65. Two of the participants are left-handed, and the others are right-handed. The participants performed the three exercises on two systems (System 1 and System 2). The participants were asked to answer survey questions and relax after every exercise. The order of the systems was random, and the participants were not notified about the difference between the systems. Both systems use the fatigue detection algorithm; however, the participants are still required to report when they feel fatigued. In this study, the participants are instructed to keep holding the end-effector for another 30-seconds, if they can, after they report fatigue. If the participants report fatigue before the algorithm detects fatigue, this extra 30-seconds allows us to calculate the temporal error of our fatigue detection algorithm. System 1 and System 2 use the fatigue detection algorithm, as mentioned above. System 1 is used primarily to evaluate the performance of the fatigue algorithm, and it does not provide any assistance to the participant when the fatigue is detected. However, System 2 has an adaptive element that changes the forces applied by the robot from resistive to assistive. In other words, the robot applies forces towards the end-point to help the participant hold the end-effector. The primary purpose of System 2 is to receive feedback about the adaptive element of the system.

2.4.3 Data Analysis

The time of subjective fatigue and the time of fatigue detected by the algorithm are analyzed to evaluate the accuracy of the fatigue detection algorithm. If the algorithm is accurate, then the difference between these time points should be small. We used the Equivalence Test, Paired Two One-Sided Tests (TOST), to check for that. The test analyzes the differences between the means of the time points and checks if they are within a particular threshold. $M_{\text{System Detection}} - M_{\text{Subjective Fatigue}}$ is the equation used to calculate the differences in the mean. TOST performs two tests. One test checks if the differences in the mean are lower than the upper threshold. On the other hand, the second test checks if the differences in the mean are higher than the lower threshold. Based on our data, we chose a threshold of ±10 seconds. Early fatigue detection is signified by the lower threshold (-10 seconds) while late detection is signified by the upper threshold (+10 seconds). Late fatigue detection in rehabilitation may lead to a dangerous accidents; therefore, it is important to know if the system recognizes fatigue late. Table 2.3 shows the TOST results, and it also includes the Mean Absolute Error.

Both p-values for the SF results are significant (< 0.001 and 0.008). That means the time difference between the subjective fatigue and the algorithm detection for SF is significantly within ± 10 seconds. Also, the mean absolute time error for SF is 9.444 seconds. For SA and EE, the results for the lower threshold test are significant, but the rest of the results are not. That indicates that the difference of the mean is significantly within the lower threshold (-10 seconds), but it could be higher than the upper threshold (+10 seconds). In addition, the mean absolute error is 12.646 seconds for SA, and 15.311 seconds for EE.

Table 2.3: Adaptive Robotic Rehabilitation: Temporal Analysis [54]. Significant results ($P < 0.05$) are marked in bold.

Exercise	P Value Upper	P Value Lower	Mean Absolute Error (s)
SF	<**0.001**	**0.008**	9.444 ± 7.803
SA	0.16	<**0.001**	12.646 ± 9.293
EE	0.709	<**0.001**	15.311 ± 9.134

2.5 Advantages, Limitations and General Observation

The solutions discussed in Section 2.3 and Section 2.4 aim to provide robot-assisted rehabilitation to patients without the continuous presence of a therapist. They are not designed to replace therapists, but to reduce the load on them. These systems can provide patients who do not have access to rehabilitation centers the opportunity to do rehabilitation exercises at home and provide the therapists with the performance data needed to evaluate the patient's progress remotely. The main limitation of these systems is that the technologies they use are still in the developing stages and are very expensive at this time. Hopefully, such rehabilitation instruments become available for a larger portion of patients in the future and have a reliable infrastructure that can easily connect patients with therapists.

CHAPTER 3

SLEEP QUALITY EFFECT ON WORK PERFORMANCE

3.1 Introduction

Sleeplessness is prevalent among professionals, and it has an association with low performance, absenteeism, and accidents at work [111]. [111] examined how work influences sleep in persons who are not at risk of sleep disorder and also how sleep influences work. They found that longer working hours are associated with reduced sleep time and that decreased sleep time is associated with more work impairments. Insomnia, obstructive sleep apnea, and shift work sleep disorder are some of the sleep disorders that might affect occupational performance. This chapter focuses on a sleep disorder called sleep apnea. [124] estimates that about 17% of adults suffer from mild to severe sleep-related breathing disorders, and it is estimated that approximately $150B loss was incurred in the United States, in 2015, due to undiagnosed apnea [123]. Sleep apnea is a sleep breathing disorder wherein the person's breathing is reduced or altogether stopped during sleep. Sleep apnea has three main types, Central Sleep Apnea (CSA), Obstructive Sleep Apnea (OSA), and Mixed. CSA happens due to dysfunction of the muscles responsible for breathing [8], whereas OSA occurs when the upper airway closes due to excessive relaxation of soft pallets in the airway [60]. Mixed sleep apnea is when both CSA and OSA occur at the same time [43]. The Continuous Positive Airway Pressure machine (CPAP) is a common way to treat OSA. CPAP passes high-pressure air through the upper airway to prevent it from collapsing and causing an apnea event. Other options include oral appliances, surgeries, behavioral, and positional therapy [107]. In particular, oral appliances adjust the position of the mandible or the tongue to stop the upper airway from collapsing. Whereas, restorative surgery would primarily target the soft palate in the airway. For instance, it would make the airway wider by removing or repositioning the excess tissue in the airway [107]. Behavioral therapy would focus on encouraging the patients to work on reducing their weight and body mass index (BMI), and positional therapy would urge the patients to sleep in a non-supine position.

3.2 Related Work

Polysomnography (PSG) is the gold standard for sleep apnea diagnosis [12]. It is an expensive, overnight study that takes place in a sleep lab. Its technical fees alone cost around $3500 (before insurance) [25]; however, that varies from a place to another. During the study night, the patients sleep in the sleep lab while a sleep-expert observes them. The study uses multiple sensors to collect data from the patients while they are asleep, including ECG, EEG, EMG, electrooculogram (EOG), oxygen saturation (SpO_2), camera, and a microphone. Since the patients need to sleep in an unfamiliar environment (i.e., sleep lab) and be connected to multiple sensors, they might be uncomfortable in their sleep, which would affect the quality of the recorded data and, in turn, underestimate or overestimate the severity of apnea. On the other hand, there are some commercial solutions to detect sleep apnea at home. Some of these solutions are ARES by SleepMed [24], and ApneaLink Air by ResMed [23]. ARES uses a headband with sensors recording SpO_2, airflow, pulse rate, snoring, and head movement and position. ApneaLink Air detects apnea by using a chest band with a device recording pulse rate, oxygen saturation, nasal flow, snoring, and respiratory effort. Moreover, some research studies also investigate solutions to diagnosis sleep apnea. For example, [14] uses a personal digital assistant (PDA) and analyzes SpO_2 data in real-time to detect apnea. Similarly, [4] uses a smartphone with other sensors for their sleep apnea detection system. They use an oximeter to measure SaO_2, an accelerometer to detect patient movements, and a microphone to record respiratory activities. They tested their system on 15 participants. Some of these participants were known to have sleep apnea. Their system's accuracy was very promising. The system could correctly identify apnea with 100% accuracy and the absence of apnea with 85.7% accuracy.

3.3 APSEN: Pre-Screening Tool for Sleep Apnea in a Home Environment

This section focuses on our apnea detection system, discussed in [55]. The system is called *APSEN*, and it stands for apnea sense. APSEN detects apnea and postural apnea using readily available and inexpensive sensors. It is an apnea pre-screening tool, and it is not a substitute for PSG. The user should consider doing a full PSG test if APSEN detects sleep apnea condition. The system provides the user with a detailed history of all apnea events and sends real-time notifications to the user in the case of severe apnea events. Also, APSEN users are expected to use the system in the comfort of their own home, without the need to visit a sleep lab.

Figure 3.1: APSEN: Nonin Oximeter Sensor [87].

Figure 3.2: APSEN: Microsoft Band 2 Smartwatch [81].

The primary devices needed for APSEN are a Bluetooth oximeter (Nonin II 9560, see Figure 3.1), a smartwatch (Microsoft Band 2, see Figure 3.2), and an IR camera (Kinect V2, see Figure 3.3). Figure 3.4 illustrates the system setup. Also, an explanation of the system's components and setup is available on an online video [1]. APSEN has two subsystems. The first subsystem is responsible for the detection of apnea events using the oximeter, and it sends apnea notifications to the smartwatch. The second subsystem detects the user's sleeping posture using the Kinect camera. By combining these subsystems, we can determine postural apnea. That can be achieved by detecting apnea events and

[1] *https://youtu.be/9pM8ZCSE8Eg*

Figure 3.3: APSEN: Microsoft Kinect V2 Camera [59].

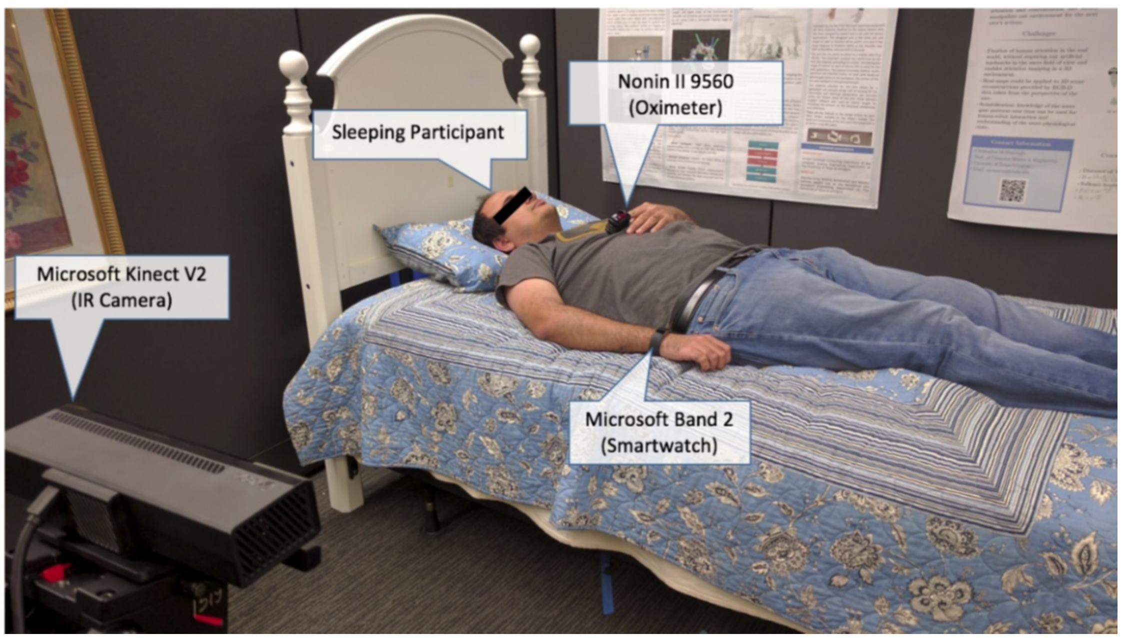


Figure 3.4: APSEN: System Setup [55].

correlating them with the user's sleeping position. These subsystems can also be used independently. The apnea detection subsystem depends primarily on the oximeter; thus, if the other devices are not available to the users, they still can do apnea pre-screening at home. Figure 3.5 shows the system architecture. The architecture shows that the smartwatch is also recording multimodal data. The smartwatch data is for experimental purposes only, and it is primarily used

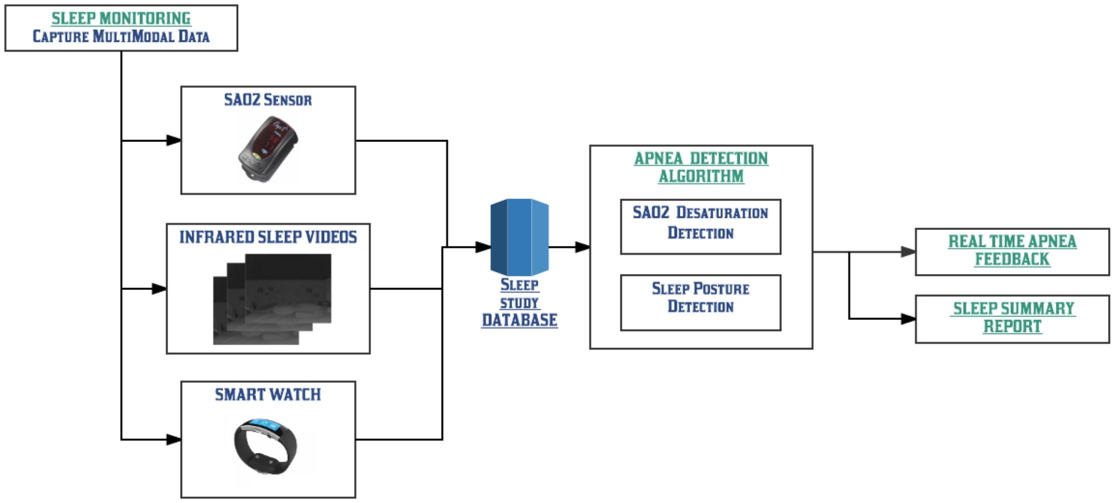


Figure 3.5: APSEN: The APSEN System Architecture [55].

to receive apnea related notifications. The rest of this section focuses mainly on the apnea detection subsystem while [55] provides the details of the posture detection subsystem.

3.3.1 Apnea Detection

SpO_2 is an indication of the oxygen level in the blood. The Nonin oximeter records SpO_2 from the user's index finger, and it sends it via Bluetooth to the main computer responsible for processing the data. Figure 3.6 shows the graphical user interface (GUI) for the apnea detection subsystem. To use the system, the users need to wear the oximeter on their left index finger and the smartwatch on their right wrist. Then, they use the GUI to connect to these two devices before they go to bed. When the study is over, the users use the GUI to disconnect these devices and see the summary of the sleep apnea results. The summary includes information about the number of apnea events, the oxygen level, the

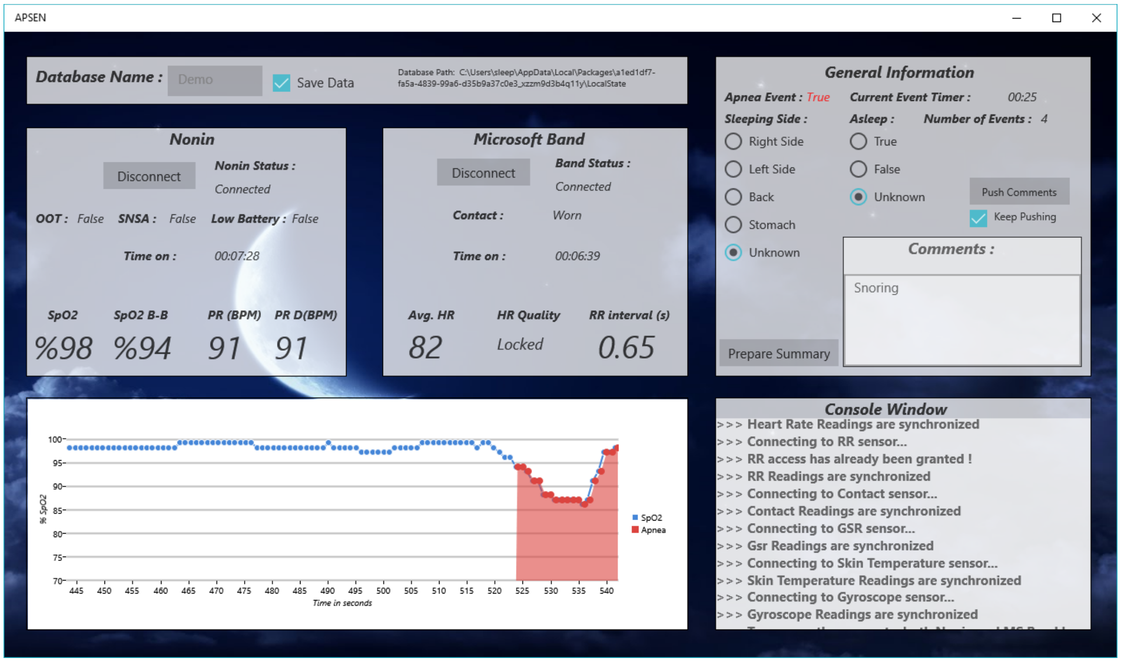


Figure 3.6: APSEN: The Apnea Detection GUI [55].

oxygen desaturation, and the length of every apnea event. It also displays an estimation of the Apnea-Hypopnea Index (AHI), which states if the apnea status is *Normal, Mild, Moderate, or Severe*. The AHI represents the average number of apnea events per hour [116]. AHI < 5, $5 \leq$ AHI < 15, $15 \leq$ AHI < 30, AHI ≥ 30 indicate *Normal, Mild, Moderate, or Severe* apnea condition, respectively. However, our calculation of AHI is an estimation because it does not count for hypopnea events.

Our apnea detection algorithm is similar to the algorithm developed in [4]. It checks if the SpO_2 desaturates. The algorithm counts the desaturation event as a possible apnea episode if it is 4% or more. If the event lasts for at least 10 seconds, the algorithm confirms that it is an apnea episode. The episode ends when SpO_2 saturates for 2 seconds or more. Figure 3.6 shows that the GUI illustrates apnea episodes by highlighting them in red on the SpO_2 graph. When the SpO_2 level drops below 80%, the apnea episode is considered dangerous. Thus, in this case, the GUI sends a notification (i.e., vibration alert) to the user's smartwatch to wake him/her up to breathe, which increases the oxygen level in the blood back to its normal level. The GUI also makes a beeping sound for a few seconds if the oximeter falls from the user's finger (e.g., due to excessive hand movement). Moreover, if the user has a family member observing the apnea pre-screening, the observing person can use the GUI to see the apnea events in real-time and also take notes about the user's sleeping behavior (e.g., the user is snoring, and sleeping on the stomach, etc.).

3.3.2 Preliminary Study

To validate our apnea detection algorithm, we tested it on data from people diagnosed with apnea and healthy people. The data is from the Apnea-ECG database [2] and it is annotated by a sleep-expert [39, 91]. The database has multiple annotated recordings, but only 8 of them include SpO_2 data, which we are using in this preliminary study. Table 3.1 shows the classification of the recordings. The dataset has four *Apnea* recordings, one *Borderline* recording, and three *Control* recordings. The dataset has one annotation per minute [3]. If the apnea episode is not happening during the start of the minute, then that minute is not marked as having an apnea episode. This annotation is different from the real-time method our algorithm uses. Therefore, we adjusted our algorithm to provide two reportings for the evaluation purpose. The first reporting is a per-minute reporting like in the database annotation, and the other reporting is the real-time reporting.

[2] $https://physionet.org/physiobank/database/apnea-ecg/$
[3] $https://physionet.org/physiobank/database/apnea-ecg/annotations.shtml$

Table 3.1: APSEN: Classification of Apnea Annotations in the *Physionet* Apnea-ECG Database [39].

Class	Description
Apnea	"Contain at least one hour with an apnea index of 10 or more, and at least 100 minutes with apnea during the recording."
Borderline	"Contain at least one hour with an apnea index of 5 or more, and between 5 and 99 minutes with apnea during the recording."
Control	"Contain fewer than 5 minutes with apnea during the recording."

Table 3.2 shows the results of our apnea detection algorithm using the two reporting methods. The 'A', 'B', 'C' letters in the table header represent the apnea classifications *Apnea, Borderline, and Control*, respectively. The accuracy of the apnea algorithm is listed in Table 3.2, Row 8. The accuracy of the algorithm is between 54.40% and 83.20% for the *Apnea* conditions, 95.70% for the *Borderline* condition, and between 99.20% and 100% for the *Control* conditions. In the *Apnea* condition, the algorithm does not have high accuracy. However, the algorithm could still indicate that the patient has a high AHI (i.e., lowest AHI for *Apnea* conditions in row 10 and 11 is 25.14), which indicates that the patient has apnea and should consider doing the PSG test. The accuracy of the algorithm in *Borderline* and *Control* is very high. These results may suggest that the algorithm can determine the absence of apnea with very high accuracy, but it has moderate accuracy in detecting severe apnea conditions.

Table 3.2: APSEN: Apnea Detection Algorithm Accuracy on the Apnea-ECG database [55].

Row	*Description*	*A01*	*A02*	*A03*	*A04*	*B01*	*C01*	*C02*	*C03*
1	Total no. of calculated apnea episodes using our algorithm (multiple annotations per minute)	586	644	427	499	6	2	10	0
2	Total no. of annotated apnea episodes from *Physionet* (1 annotation per minute)	470	420	246	453	19	0	1	0
3	Total no. of calculated apnea episodes using our algorithm (1 annotation per minute)	392	277	217	233	4	1	3	0
4	Total no. of apnea episodes matching in both *Physionet* and our algorithm (1 annotation per minute)	390 out of 470	260 out of 420	185 out of 246	231 out of 453	1 out of 19	0 out of 0	0 out of 1	0 out of 0
5	Total no. of 'no apnea' episodes from *Physionet* (1 annotation per minute)	18	107	272	38	467	483	500	453
6	Total no. of calculated 'no apnea' episodes using our algorithm (1 annotation per minute)	96	250	301	258	482	482	498	453
7	Total no. of 'no apnea' episodes matching in both *Physionet* and our algorithm (1 annotation per minute)	16 out of 18	90 out of 107	240 out of 272	36 out of 38	464 out of 467	482 out of 483	497 out of 500	453 out of 453
8	Our algorithm accuracy (1 annotation per minute)	83.20 %	66.41 %	82.05 %	54.40 %	95.70 %	99.80 %	99.20 %	100 %
9	Calculated AHI from *Physionet* annotations (1 annotation per minute)	57.79	47.82	28.49	55.36	2.35	0.00	0.12	0.00
10	Calculated AHI from our algorithm (1 annotation per minute)	48.20	31.54	25.14	28.47	0.49	0.12	0.36	0.00
11	Calculated AHI from our algorithm (multiple annotations per minute)	72.05	73.32	49.46	60.98	0.74	0.25	1.20	0.00

3.3.3 Overnight Study

Ten participants were recruited for the overnight study. The participants were healthy male students and university staff between the age of 24 and 31 years old. The study took place in a simulated apartment setup, as shown in Figure 3.4. The participants came to the study around their bedtime. They started the study by signing the consent form and wearing the oximeter and the smartwatch. Before they went to bed, they interacted with the GUI to connect the sensors. The participants were instructed to not use a blanket because the blanket would cover the participant's body and affect the detection of sleep-posture. However, the study was recorded in the dark since the posture detection algorithm used IR images instead of color images. During the experiment, one of the study personnel was available at the side of the room, using the GUI, to take notes about the participant's sleep behavior, mark the ground truth of the sleep posture, and check if the oximeter is attached to participant's finger. The study produced 55 hours of data recordings. During the study, the participants slept between 2.62 and 7.89 hours, with a mean of 5.5 hours. The estimated AHI of the participants was between 0 and 1.88, with an average AHI of 0.29. That confirms that the participants were healthy and had a normal condition, AHI < 5.

3.4 Advantages, Limitations and General Observation

This chapter discussed the relationship between sleep quality and work performance briefly, and it presented our apnea pre-screen system. The system was divided into an apnea detection subsystem using SpO_2, and a posture detection subsystem using IR images. The results of the preliminary study showed that the apnea detection algorithm produced moderate results. The main limitation of the study is that participants are not allowed to use a blanket, which could impact their sleep quality. Also, the relationship between apnea and BMI of the participants was not investigated in this study. High BMI (i.e., obesity) is viewed as a factor in the pathogenesis of sleep apnea [32]. We hope that the system is evaluated in the future with people diagnosed with apnea and people with various BMI.

CHAPTER 4

BRAIN-COMPUTER INTERFACES TO UNDERSTAND MENTAL STATE

4.1 Introduction

Brain-Computer Interface (BCI) describes the neurofeedback interface that takes brain data and processes into a different useful form. The focus of BCI application has expanded widely in the last two decades. BCI expanded from applications diagnosing mental disorders to applications facilitating our daily life activities. For example, BCI can help people meditate and relax, and it can also offer people with disabilities an unconventional way to control augmentative technologies. By having access to brain data, BCI can process these data to estimate human feelings (e.g., sadness and happiness), improve human-computer interaction, and make it adaptive [94].

EEG is a measure of the electrical activity resulted due to brain activity (i.e., cortical activity) throughout the scalp [1]. EEG patterns reflect various phenomena that are investigated in various research fields. For instance, EEG has been used in memory research [42] and sleep studies [15]. It is also used in conjunction with other brain imaging techniques (e.g., functional magnetic resonance imaging, etc.) to study abnormal brain activities like seizures and stroke [97, 102]. On the other hand, other studies use EEG to measure mood states. Research studies usually segment the EEG signal into five frequencies. These frequencies are delta (Δ, 1-4 Hz), theta (Θ, 5-8 Hz), alpha (A, 9-13 Hz), beta (B, 12-30 Hz), and gamma (Γ, 30-50Hz) [97]. Studies have found that increased cortical alpha and beta activity can be an indication of stress [5, 68, 69]. In addition, frontal alpha activity and increased theta activity, but no change in frontal beta activity can reflect engaged enjoyment [46, 61, 95].

Conventional EEG devices used in laboratories usually have multiple electrodes (i.e., 32 or more) that record data from many brain regions. The electrodes are placed on the scalp using special get or paste, which provides reliable connectivity and less noise. Therefore, EEG studies often take long-time to set up. After the data is collected, the data analysis requires advanced techniques and statistical knowledge. All of that makes EEG studies expensive and challenging to set up, limiting their use in daily life activities. In recent years, there are have been several EEG devices that work on making EEG accessible to more applications and users, leading to an expansion of EEG usage into everyday life applications and outside of the laboratory use. For example, some of these devices use fewer electrodes (i.e., 2 to 16 electrodes), focus on specific brain regains, and use dry electrodes that are easier to use and clean. Some of them also come with computer and or phone programs that have pre-made algorithms to clean, process, and classify the EEG data (e.g., concentrated, engaged, relaxed, etc.).

4.2 Lifestyle EEG Instruments and Applications

Nowadays, many compact EEG devices are available for regular consumers and intended for use in daily life applications. Some of the companies that produce such devices are InteraXon, Emotiv, Melon, NeuroSky, and Versus. Many of these devices are affordable, wireless and easy to set up. Also, there are some open-source platforms (e.g., OpenViBE, OpenEEG, BCI2000, and MuLES) that remove some of the burden of using EEG. They make it easier for researchers to take advantage of the information EEG signals provide. Such efforts helped facilitate the usage of EEG in both novel and established fields [16]. For instance, [62] has used open-source hardware to detect driver drowsiness from EEG data and to alert the driver when it is unsafe for him/her to drive. In [57], the researchers used Muse EEG headband to detect cold-induced pain. Similarly, [63] analyzed EEG signals to measure concentration and relaxation. Another interesting work mimicked the concept of "light bulb/idea" metaphor [73]. They mounted a light bulb on a helmet and programmed it to turn on when the EEG data reflect an increase in thinking and focus. Based on this new scope of EEG applications, we have designed a system [1] that utilizes EEG data from a consumer-grade device to measure user enjoyment.

4.3 Brain-EE: Brain Enjoyment Evaluation using Commercial EEG Headband

This section focuses on our Brain-EE system [1]. Brain-EE analyzes EEG signals to measure user's engaged enjoyment and predicts their activity preference.

4.3.1 Equipment and Software

Hardware

Muse headband is the EEG device used in this study. It is a consumer-grade, low-cost EEG headband. Multiple studies were able to achieve good results with it [16, 36, 57, 63, 73]. It collects data from four channels, and it has three

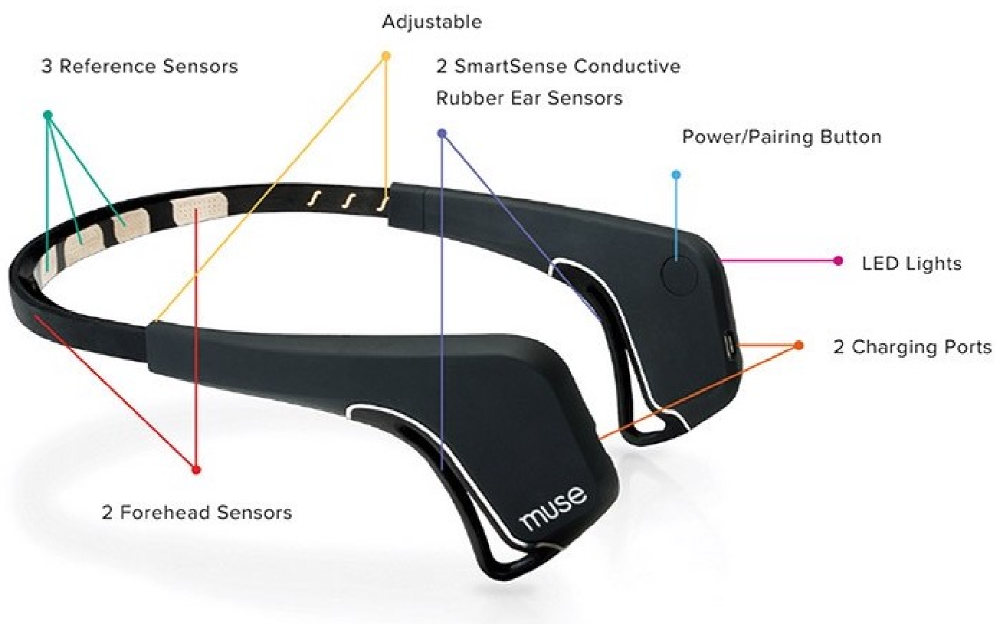


Figure 4.1: Brain-EE: Muse Headband [83].

reference electrodes. It also collects 3-axis accelerometer data. Figure 4.1 shows the Muse device and highlights its main components. The data is collected from the four EEG electrodes. Two of the electrodes are located on the forehead while other two are located above the ears, as shown in blue in Figure 4.2. We collect the Muse data via Bluetooth, and the data is sent to a computer for processing. Also, we are using an Android tablet to play white-noise sound and to allow the participants to play the games used for this study.

Software

In this study, we try to predict the user's preference between two games. Specifically, we used 'Soccer 2016' and 'Piano Tiles 2'. The users play the games on an Android tablet while we record their EEG data with the Muse headband. To access the Muse data, we are using the Muse SDK. We designed a GUI using

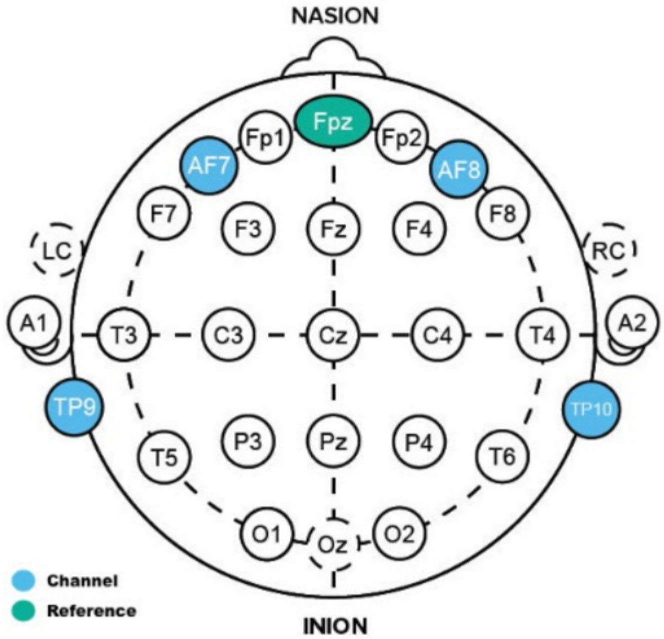


Figure 4.2: Brain-EE: Muse electrode locations by 10-20 International Standards [83].

MATLAB to record, filter, process, and analyze the EEG data. The GUI collects EEG data in intervals of 120 seconds. It does that twice. In particular, the system records EEG data for 2 minutes while the user is playing the first game, and then it records another 2 minutes of data while the user is playing the second game. After that, the system analyzes the data and predicts which game the user preferred the most.

4.3.2 User Study

Fifteen people (13 males, 2 females) participated in the study. The participants were between the age of 20 and 35 years old. The study personnel explained the experiment protocol verbally to the participants and answered any question

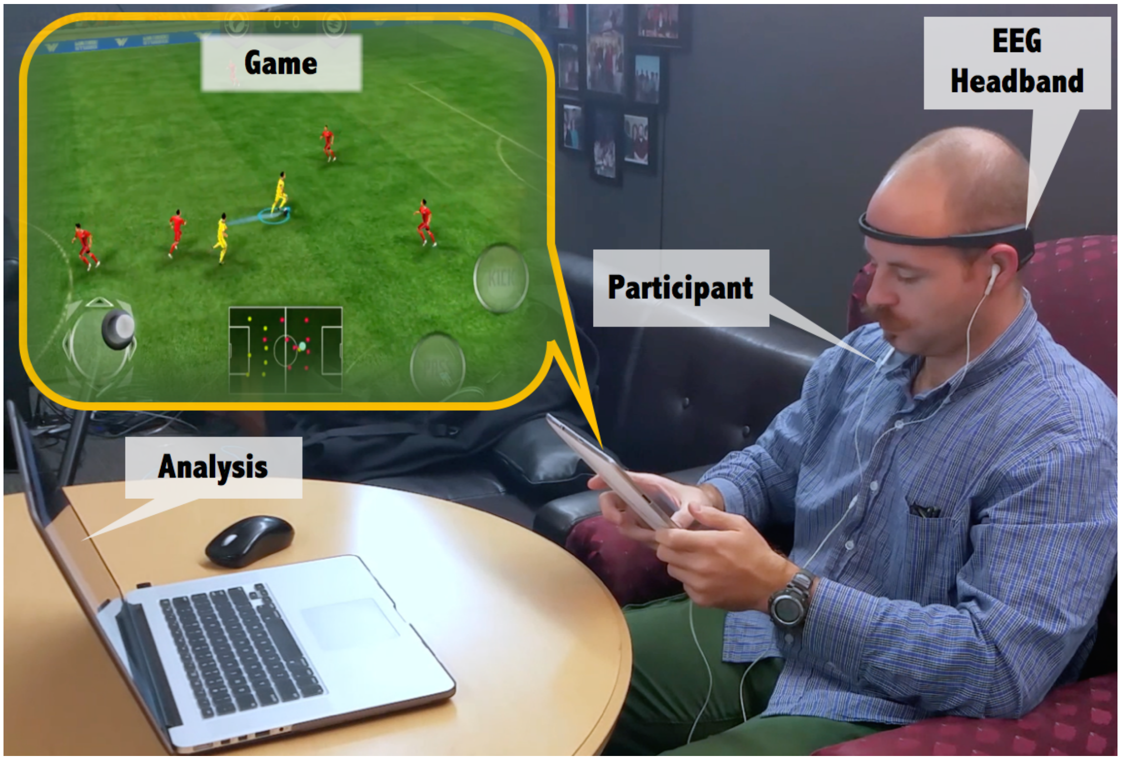


Figure 4.3: Brain-EE: System Overview [1].

they had. Figure 4.3 illustrates the experiment setup. The participants then signed the consent form and started the study. First, the study personnel helped the participants wear the Muse headband and made sure that all the electrodes are connected well to the user's skin. Then, the participants performed the three stages of the study (i.e., 'Relax', 'Game 1' and 'Game 2'). Each stage is 2 minutes long. The first stage is the baseline stage, where the user listens to white-noise sound and relaxes. In the next two phases, the participants play the games (i.e., 'Soccer 2016' and 'Piano Tiles 2') for 120 seconds. The participants were asked to answer a general mood survey after every game. This online video [1] explains the experiment protocol in details.

[1] *https : //youtu.be/brFZ93Omq5U*

4.3.3 Results and Discussion

EEG signals vary from a person to another, which makes it challenging to find a classification solution that uses EEG signals and can easily generalize to a large population [57]. However, the goal of this study it to generalize our enjoyment classifier to work with varies people and activities. Thus, we focused on comparing user's data while doing two activities, rather than comparing the data of two people. It is simpler to compare two datasets from the same source than from two distinct sources. To achieve our goal, we used the t-test to check the difference between the user's two datasets, and we used the results of the t-test into a machine learning algorithm for classification. This approach is much simpler than some of other known methods. For instance, one of the common methods (used in Chapter 5) extracts features from the signals and use these features to train machine learning algorithms.

To prepare the data for the t-test, we averaged signal from the four channels. This was done for all the 5 frequency bands. That resulted in 5 arrays of data per stage (i.e., 'Relax', 'Game 1', and 'Game 2'). Then, we used the t-test to compare 'Game 1' data and 'Relax' data, and to compare 'Game 2' data and 'Relax' data. Then, we compared the differences between the results of the two t-tests. After that, the t-test results were used to train a linear regression classifier. We used the data of 10 participants for training and data of 5 participants for testing. Equation 4.1 [1] shows the tuned linear regression equation. The y value predicts the enjoyment level. Thus, the system predicts that the game resulting in the higher y value is the game the user enjoyed the most. When the algorithm was tested on the test data, it gave 100% accuracy. However, when algorithm was used to detect enjoyment as a 'Yes' or 'No' problem, instead of which game was enjoyed more, the accuracy dropped to 60%.

$$y = -0.0651 - 0.0136\Delta + 0.0256\Theta - 0.0072A + 0.0009B - 0.0032\Gamma \qquad (4.1)$$

Table 4.1: Summary of the t-test averages across all the frequency bands [1]. The coefficients belong to Eq.4.1.

t-test of Frequency	Δ	Θ	A	B	Γ
Coeff.(C)	-0.0136	0.0256	-0.0072	0.0009	-0.0032
Enjoy (E)	34.49	60.97	29.33	64.01	23.01
No Enjoy (NE)	1.73	16.26	5.54	47.67	28.14
$C \times E$	-0.4691	1.5608	-0.2112	0.0576	-0.0736
$C \times NE$	-0.0235	0.4162	-0.0399	0.0429	-0.0900

We mentioned earlier that frontal alpha activity and increase in theta, but no change in frontal beta activity, has been correlated with engaged enjoyment [46, 61, 95]. Interestingly, our findings partially agree with that. Table 4.1 presents the coefficients of the linear regression classifier (i.e., Equation 4.1) and their multiplication with the average t-test results of the enjoyed and not enjoyed game (i.e., or less enjoyed game). We can see in Table 4.1 that theta is higher in '$C \times E$' (i.e., higher in the enjoyed game) than in '$C \times NE$' and its coefficient is positive. That might be an indication that an increase in theta is a sign of more enjoyment. On the other hand, alpha and delta values are similar in '$C \times E$' and '$C \times NE$', and have negative coefficients. That might indicate that alpha and delta do not strongly correlate with enjoyment. However, we need to note that these assumptions are made based on a small dataset. More data is required to be able to draw a concrete conclusion. Also, [1] provides a more detailed explanation of the analysis.

4.4 Advantages, Limitations and General Observation

This chapter discussed briefly brain-computer interfaces and their uses in general. It also focused on consumer-grade EEG devices, and how we utilized Muse EEG head to detect engaged enjoyment. We mentioned that consumer-grade EEG devices provide promising results, and they are easier to use and set up compared to the traditional EEG devices used in extensive research studies. However, we need to clarify that EEG data in general, and from devices with dry electrodes in particular, has a lot noise and require extensive data processing. For example, body movement, jaw movement, and blinks are reflected as noise in the EEG data. In addition, we noticed in several EEG studies that the participants often feel irritated and uncomfortable after wearing the EEG headset for a long time. Therefore, we believe there should be more research on making EEG devices more comfortable, practical, and accurate.

CHAPTER 5

9PM: A MULTIMODAL COGNITIVE AND PHYSICAL ASSESSMENT PLATFORM

5.1 Introduction

Human impulsive behavior and physiological responses might tell a lot about one's cognitive and physical state and performance. To confirm that, we designed a tool to detect the user's physical and cognitive performance from their task performance, behavioral, and physiological data. We also studied the correlation between physical and cognitive performance. This project's name is '9PM', which stands for 9-Peg Moves. 9PM is a novel system that combines cognitive and physical assessment in a single platform, and it utilizes statistical analysis, machine learning, and multimodal data for assessment. The tasks used in this project are based on the principles of standard physical and cognitive tests. 9PM combines a modified version of the 9-Hole Peg Test (9-HPT, a manual dexterity test) [121] with cognitive tests. These cognitive tests are based on principles of the Stroop Test [109], the Wisconsin Card Sorting Test (WCST) [41], and the NIH Toolbox Picture Sequence Memory Test (PSMT) [30]. The principles of these tests are used to create five tasks that users perform in random order. The tasks are designed to allow the system to precisely record all user moves and calculate the time and correctness of these moves.

5.2 Physical and Cognitive Assessment

5.2.1 Physical Tasks

The 9-HPT is a quantitative, timed test that is considered as the gold-standard test to assess manual dexterity [99]. This manual dexterity test is commonly used in assessing children, adults, and patients with Multiple Sclerosis (MS) and Parkinson's disease [96, 74, 35, 31]. In the 9-HPT, users transfer nine pegs, one at a time, from a shallow container to another container with nine holes, and then back to the shallow container [84]. They do that twice with the dominant hand and twice with the non-dominant hand while being timed with a stop-watch. Figure 5.1 illustrates a 3D model of the original 9-HPT. [74] provides the dimensions and more design details. For the 9PM system, we modified the 9-HPT board, and we created a novel test board, called *9PM board*. The 9PM board is computerized and allows us to combine this physical test with cognitive tests, and track user actions. It has three colored areas, and each area has nine holes, as shown in Figure 5.2. The blue and red areas are the start areas, and the white area is the destination area. The user is required to take action (i.e., move a peg from the blue or red area) based on the instructions on the screen. These instructions are based on cognitive tests. In contrast to the original 9-HPT, 9PM users receive an audio-feedback for correct/wrong moves.

5.2.2 Cognitive Tasks

The cognitive tasks used in 9PM are based on the Stroop test, WCST, and PSMT. The Stroop test reflects on one's cognitive capacity to maintain a specific course

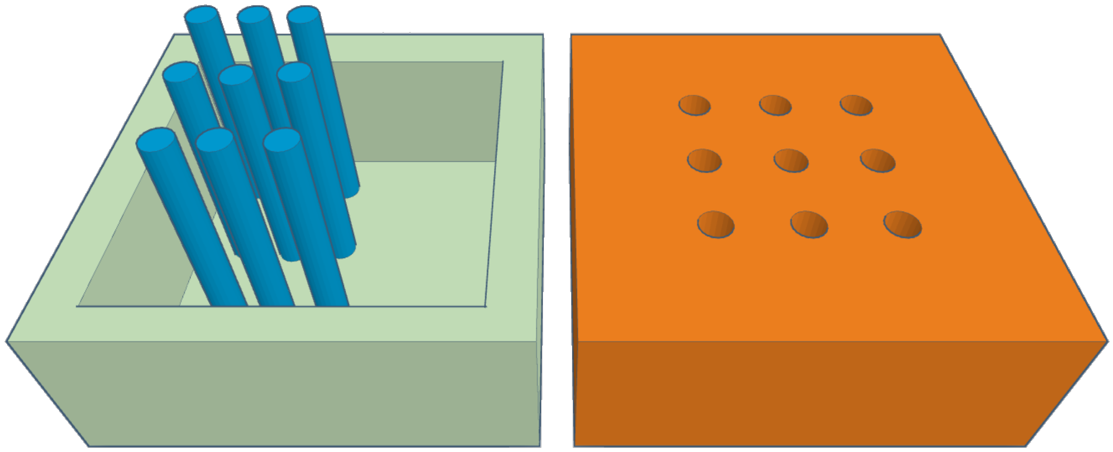


Figure 5.1: 9PM: 3D Model of the Original 9-HPT

Figure 5.2: 9PM: 9PM Board, a Novel Modified Version of the 9-HPT

of action in the presence of an interfering stimulus [26]. For example, the user is required to name the font-color of a word rather than reading it. Based on the Relative Speed of Processing theory, the process of reading words is faster than calling their colors [72], as reading words is an automatic process that requires much less attention than naming colors. That can mean when the user tries to name the font-color of a word, the text of the word is processed faster by the brain, and it interferes with the process of calling the font-color, which could lead to a slower reaction.

9PM uses two variations of the Stroop test. In the first variation of the Stroop test (*Stroop-Text*), the stimulus is displayed in black font-color (i.e., 'Red' or 'Blue'), and the user needs to respond based on the semantic meaning of the stimulus. This task can be an indication of the user's cognitive processing speed [29, 50]. In the second variation (*Stroop-Color*), the semantic meaning of the stimulus and the font-color might be similar or conflicting (i.e., 'Red', 'Red', 'Blue', or 'Blue'), and the user is required to name the font-color. This variation can give an insight into the user's inhibitory control ability [104].

In WCST, the user needs to match cards based on the shape, color, or the count of items drawn on the cards. However, the user is not told how to match them and needs to figure it out initially based on trial and error, and the matching rule changes randomly after a few rounds. So, the WCST assesses the user's abstract reasoning and task-shifting ability [9]. The 9PM version of the WCST has two matching rules: the semantic meaning of the stimulus or its font-color. Since the 9PM version of the WCST is similar to the WCST and the Stroop test, we decided to call it *Stroop-Shifting*. The stimulus in *Stroop-Shifting* is either 'Red', 'Red', 'Blue', or 'Blue', and the user is required to shift between responding to font-color and the text of stimulus. The requirement (text or font-color) randomly changes after N moves, where N is a random number between 3 and 9. For instance, the user is provided with a stimulus 'Red' and is required to determine whether he/she should respond to the font-color or the text of the stimulus. The matching rule will stay the same for at least three moves. For example, if the user acts based on the text of the stimulus, the system would provide audio-feedback to illustrate whether the selection was correct or wrong. If the user hears the wrong-buzzer, he/she knows that in the next two moves, at least, they should respond to the font-color of the stimulus rather than its text.

The PSMT is a cognitive test to examine episodic memory. Episodic memory "is one of the most important cognitive domains that involves acquiring, storing and recalling new information" [30]. In the PSMT, the user is presented with a sequence of pictures and verbal descriptions, and the user should memorize and recall the sequence. The 9PM version of the PSMT is called *Color-Sequence*. *Color-Sequence* has been inspired by the PSMT and another task called *Sequence Learning* [119]. *Color-Sequence* is a working memory and sequencing task where the user is asked to remember and repeat a sequence of colors. The sequence will have a combination of the stimuli 'Red' and 'Blue'. An example of the sequence is "Red Blue Blue Blue Red Red Blue Red Blue". Each stimulus is presented to the user for 3-seconds before it disappears and the next stimulus appears. The user needs to memorize the 9-stimuli and recall them after that.

5.2.3 9PM Tasks

9PM utilized the above-mentioned physical and cognitive tasks to produce five tasks. Table 5.1 summarizes these tasks. In all the tasks, the participants are asked to use their dominant hand to move pegs on the 9PM board, one at a time. Also, they are instructed to work as quickly as possible and to avoid making any wrong moves. If they hear the wrong-buzzer audio-feedback, they need to move on without making any corrections. To reduce the load on the participants, they have the freedom to pick/place any of the 9-pegs in a certain area without a specific order. For instance, if the participant is asked to move a peg from the blue area to the white area, he/she can pick the middle peg from the blue area and place it into the hole in the right-bottom corner in the white area.

In Task 1 (T1), the participant uses the 9PM board to make 72 peg-moves in a single round without a break. Right-handed participants are instructed to move the pegs between blue (closer to the right-hand, see Figure 5.2) and the white area. Similarly, left-handed participants are requested to move the pegs between the red and the white area. To make the 72-moves, the participants start with 9-peg moves from one of the start areas (i.e., either blue or red area) to the white area (destination area). Then, they need to transfer them back to the start area, which counts as another 9 moves. They need to repeat this process four times to complete the 72 movements.

Task 2 (T2) combines the physical task of moving the pegs on the 9PM board with the Stroop-Text task. The participant performs 4 rounds of this task, and every round has 9 moves. Between the rounds, the participant answers a short survey (Appendix A.1.2). As explained in Section 5.2.2, a stimulus (i.e., 'Red' or

Table 5.1: 9PM: Tasks Summary

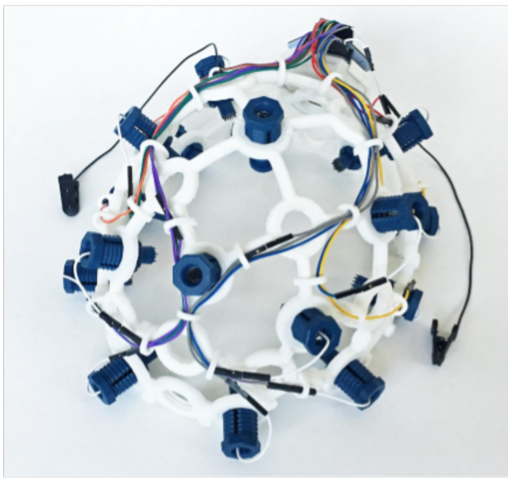
Task	Measure	Physical or Cognitive	Description
Baseline	Baseline	None	- Relax for 3-minutes with closed eyes.
T1	Hand Movements	Physical	- 1 Round (1 round = 72 moves). Move the 9-pegs one at a time from one start area to the destination area. Then, move them back one at a time to the start area. Repeat 4 times without a break.
T2	Cognitive Processing Speed	Both	- 4 Rounds (1 round = 9 Moves). Move the 9-pegs one at a time from the start areas to the destination area based on the semantic meaning of the stimuli.
T3	Inhibition Control	Both	- 4 Rounds (1 round = 9 Moves). Move the 9-pegs one at a time from the start areas to the destination area based on stimuli font-color.
T4	Cognitive Shifting Flexibility	Both	- 4 Rounds (1 round = 9 Moves). Move the 9-pegs one at a time from the start areas to the destination area based on either the semantic meaning or the font-color of the stimuli.
T5	Working Memory	Both	- 4 Rounds (1 round = 9 Moves). 9 stimuli displayed on the screen for 3 seconds, one at a time. Memorize the 9-stimuli, and then move the 9-pegs one at a time from the start areas to the destination area based on the stimuli (semantic meaning agrees with font-color).

'Blue') is presented on the screen, and the participant is required to move one peg based on the semantic meaning of the stimulus. For instance, the participant would need to transfer a peg from the blue to the white area if the stimulus were 'Blue'. The system provides audio-feedback (correct/wrong buzzer) when a peg is picked-up and when it is placed-in. The next stimulus will be shown immediately after the participant completes the transfer of the peg to the white area. After the participant moves 9 pegs and finishes the round, the system stops recording data, and the study personnel or the participant moves the pegs back to the start areas.

Task 3 (T3) and Task 4 (T4) are very similar to T2, except that the stimuli font-color is red or blue, and that the participant needs to respond to them differently. T3 and T4 also have 4 rounds, with 9 moves per round. T3 combines the physical task of moving the pegs on the 9PM board with the Stroop-Color task. This time, the participant needs to act based on the font-color of the stimulus, rather than its semantic meaning. For instance, if the stimulus were 'Red', the participant would need to move one peg from the blue area to the white area. That is because the font-color is blue. However, T4 combines the physical task of moving the pegs on the 9PM board with the Stroop-Shift task. For the first stimulus, the participants do not know whether they have to follow the semantic meaning or the font-color. They figure that out after they pick-up a peg and hear the audio-feedback. The rule (semantic meaning or font-color) changes every 3-9 moves, and it does not carry-on to the next round.

Task 5 (T5) also has 4 rounds (9 moves per round), but it has a different structure. It combines the physical task with the Color-Sequence task. Thus, the participants observe and memorize 9 stimuli before they start moving 9-pegs.

Each stimuli (i.e., 'Red' or 'Blue') appears in the middle of screen for 3-seconds before it disappears and the next one is shown. After all the 9-stimuli have been displayed, the participant then starts moving 9-pegs from the corresponding areas to the white area.



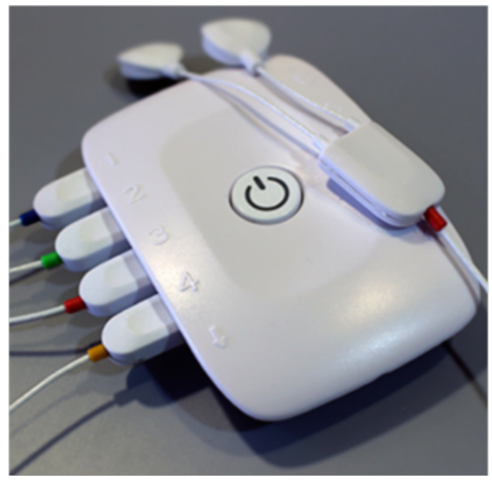


Figure 5.3: 9PM: Physiological Devices: OpenBCI ULTRACORTEX MARK IV sensor (left) [89], and Biosignalsplux Explorer unit (right) [11].

5.3 Data Collection

Physiological, behavioral, and performance data are the three main types of data collected in this project. Other related data, such as task name, user surveys, and demographic details (i.e., age, gender, and handedness) are also recorded.

5.3.1 Physiological Data

EEG, EDA, and ECG are the three physiological data that are collected. EEG data is obtained using the OpenBCI ULTRACORTEX MARK IV sensor [89], whereas EDA and ECG data are collected using the Biosignalsplux Explorer unit [11]. Figure 5.3 shows both of the devices.

5.3.2 Behavioral Data

The behavioral data collected are user surveys, hand movements, and facial expressions. Subjective behavioral data such as mental and physical workload, sleepiness, difficulty concentration, enjoyment, interest, distress, and attention are gathered from user surveys. The user's hand movement data is obtained using an IMU sensor strapped on the wrist of the user's dominant hand. The 9-axis IMU sensor used is the MetaMotionR sensor from Mbient Lab [80], Figure 5.4. The pattern of hand movements might be a useful indicator of stress and mental and physical workload. Also, a web-camera is used to record the user's facial expressions throughout the experiment; however, the web-camera data was collected for experimental purposes, and it is not part of the analysis in this dissertation.

5.3.3 Performance Data

The performance data include task-score, move-time, and reaction-time. The task-score is the percentage of correct movements. The move-time is the elapsed time from taking out a peg from the start area to placing it in the destination

Figure 5.4: 9PM: Mbient Lab MetaMotionR 9-axis IMU Device [80].

area, which might reflect on the user's physical ability and performance. The reaction-time is the time the user takes to decide his/her next move, which can reflect on the user's cognitive strength and performance. In the tasks where the user needs to observe the stimulus on the screen immediately before deciding on the next move (i.e., T2, T3, and T4), the reaction-time is the elapsed time from the appearance of the stimulus on the screen to the time the user picks up the peg from the start area. In other tasks (i.e., T1 and T5), the user is given the stimuli/stimulus in advance; thus, the reaction-time is calculated as the elapsed time between placing the peg in the destination area and picking up the next peg.

5.4 Experiment Setup

5.4.1 System Architecture

The system consists of 7 devices. The first device is the main computer that stores all the data and provides audio feedback to the user. The main com-

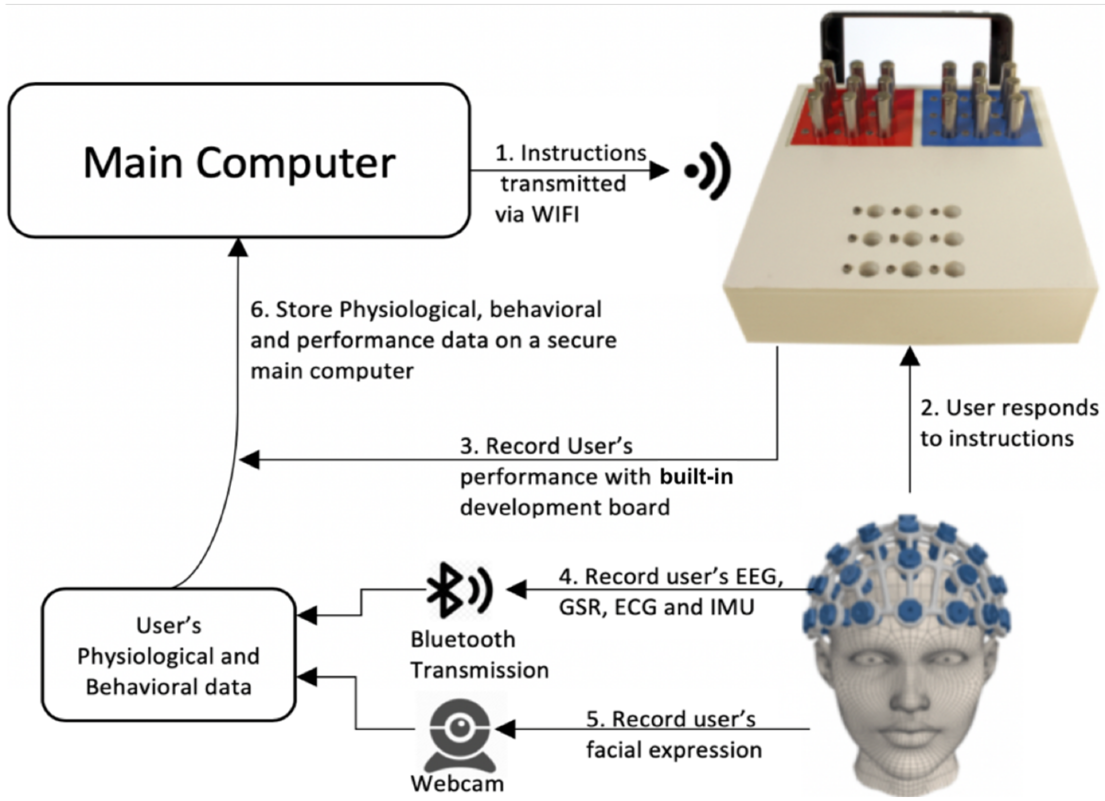


Figure 5.5: 9PM: System Architecture.

puter sends instructions through a virtual server to a specific public IP address. This method allows the displaying device to be any device that can run a web browser (e.g., smartphone, tablet, regular computer, etc.). The displaying device in Figure 5.5 is a smartphone running Chrome web-browser and displaying the instructions/stimuli on its screen. The user sees the instructions and responses to them by moving the pegs. The holes in the 9PM board act like switches that turn off or on when the user picks up or places the pegs. All the 27 holes are connected to a Teensy 3.5 development board embedded inside the 9PM board. The development board records all user moves and sends them in realtime via a USB cable to the main computer. Meanwhile, OpenBCI EEG sensor, Biosignal Plux (ECG and EDA) unit, the IMU device, and a web-camera are recording the user's physiological and behavioral data and transmitting them to the main computer. Figure 5.5 illustrates an overview of the system architecture. The system is designed to be modular, and the devices can be replaced.

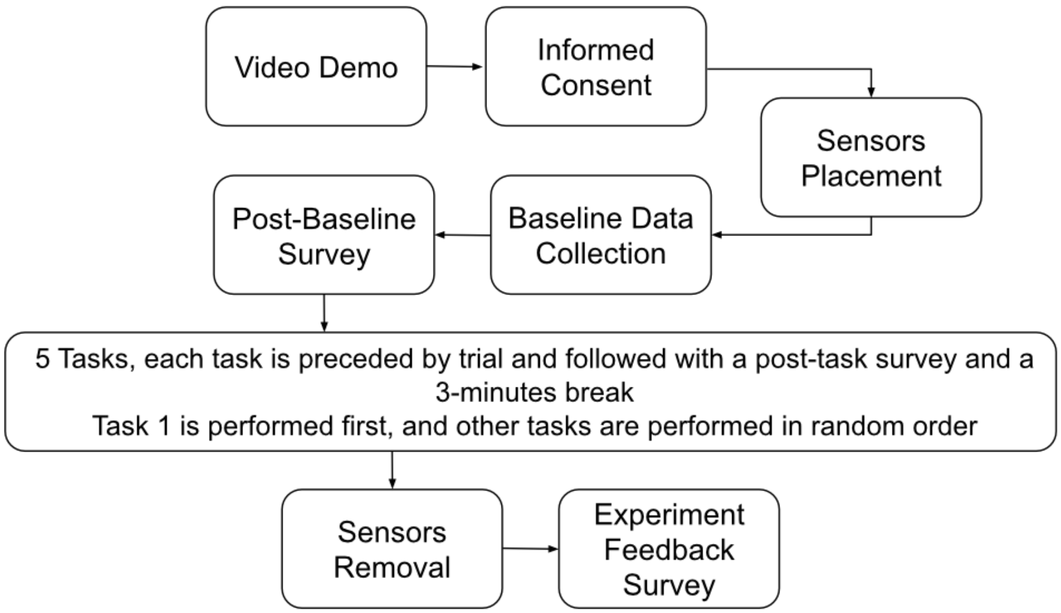


Figure 5.6: 9PM: Experiment Workflow.

5.4.2 Experiment Workflow

The study consists of a baseline data collection stage and a set of 5 tasks, as illustrated in Table 5.1. Figure 5.6 summarizes the experiment workflow. At the beginning of the experiment, the participant is asked to sit on a chair at the experiment desk and watch a video[1]. This video is a brief demo on the experimental procedure and on how to wear the sensors. After that, the participant is asked to read and sign the consent form. During this process, the study per-

[1] $https://youtu.be/1O5pmqFOFFQ$

sonnel is available to answer the participant's questions, if any. Next, the participant wears the sensors in private. The sensors that the participant wears are Biosignalsplux Explorer unit, OpenBCI ULTRACORTEX MARK IV sensor, and the MetaMotionR sensor, as explained in Section 5.3. Figure 5.7 shows the experiment setup and sensors placement. After that, the study starts with a baseline stage. In the baseline stage, the system records sensors data while the participants are closing their eyes and relaxing on a chair for 3-minutes. Then, the participants are asked to participate in a survey (Appendix A.1.1). The purpose of this survey is to collect the participants' baseline subjective data (e.g., they feel sleepy and have difficulty concentrating before doing any of the 5-tasks).

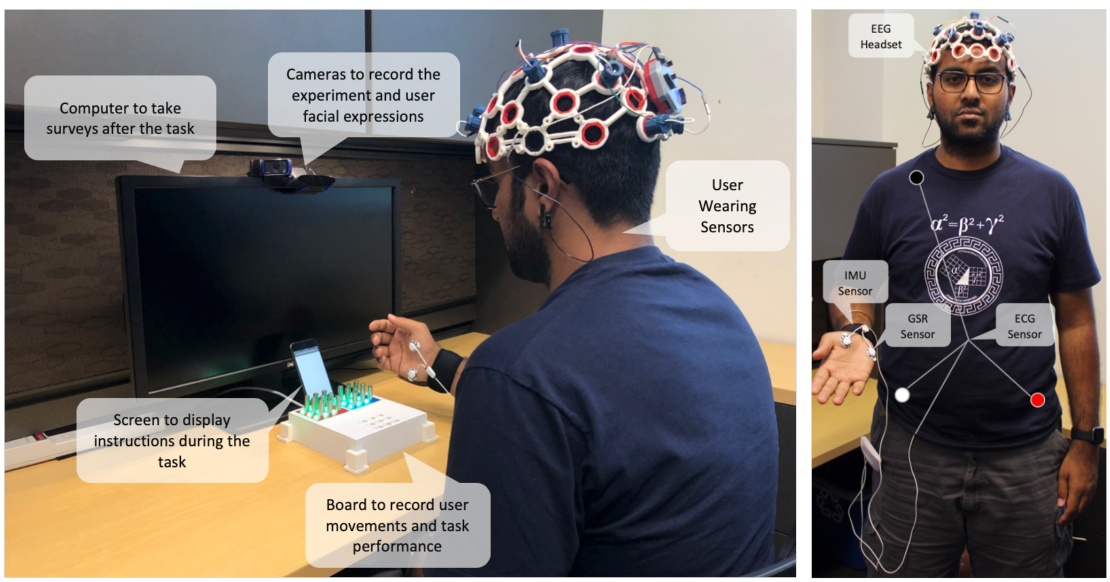


Figure 5.7: 9PM: Experiment Setup (Left) and Sensors Placement (Right).

The participants then perform the 5-tasks. They start with T1, and then they do the other tasks in random order (e.g., T1, T5, T2, T4, T3), with a 3-minutes break between them. Performing the tasks in random order might limit the cascading-effect of fatigue and experience effect, which could jeopardize the quality and reliability of the data. The cascading-effect of fatigue is when the participants feel tired and perform poorly in the final stages. In contrast, the experience-effect is when the participants get used to the task and perform much better in the final stages. To make sure that the participants understand the task rules fully, they are asked to do a trial before each task. T1 trial is just an 18-moves version of the task; however, the trial of the other tasks is the same as a regular round. After every round/task, the participants are asked to fill out a survey. They fill out a short survey (Appendix A.1.2) after each of the first three rounds in T2, T3, T4, and T5. However, after T1 and the 4^{th} round of the other tasks, they fill out a lengthy survey (Appendix A.1.3). Finally, once they finish all the tasks, they remove the sensors in private, and then they complete a final feedback survey, Appendix A.1.4.

5.5 Research Questions and Hypotheses

This study focuses on answering three research questions about physical and cognitive performance. Based on the design of the study, we have four hypotheses. Section 5.5.1 and 5.5.2 list the research questions and hypotheses, and they are discussed in details in Section 5.7.

5.5.1 Research Questions

Q1: Can the user's cognitive and physical performance be detected using: physiological data, behavioral data, and or performance data?

Q2: Which data give a better indication of user's cognitive and physical performance?

Q3: How cognitive and physical performance correlate?

5.5.2 Hypotheses

H_1 : **Difficulty would increase with the task level**: Each task would be cognitively more difficult and challenging than the previous task level. T1 would be the easiest, and T5 would be the most difficult. We assume that each task level is more cognitively demanding than the task level before it; thus, we expect the difficulty to increase and the performance to decrease.

H_2 : **There would be a significant difference in the reaction-time between the tasks**: We anticipate that the participant would not need a long-time to react to easy tasks, so we expect the reaction-time to be significantly shorter in easier tasks.

H_3 : **There would not be a significant difference in the move-time between the tasks**: Since the physical aspect of the tasks is the same in all them, we expect that the move-time to be similar in all of them.

H_4 : **Participant's physiological, behavior, and performance data would be a good indicator of the participant's cognitive and physical workload and performance**: We expect that we would be able to design machine learning algorithms that can use participants' data to detect their current level of the physical and cognitive workload.

5.6 User Study, Results and Analysis

To validate our hypotheses, we conducted a user study, and we analyzed the data. The code and data used for this study can be found online [2]. Section 5.6.1 explains the demographics of the participants and the number of rounds and surveys they performed. The analysis of the data is divided into three parts. The first part (discussed in Section 5.6.2) focuses on analyzing the subjective survey data to show how the participants felt after the tasks. Then, Section 5.6.3 shows the performance data analysis, which gives an insight into how the participants performed in different tasks. For example, it looks into which aspect of performance (physical, cognitive, or both) is effected in more challenging tasks. Finally, Section 5.6.4 discusses how we utilized machine learning and multimodal data in assessing the participant's mental and cognitive state/performance (e.g., physical tiredness, distress, sleepiness, etc.).

5.6.1 User Study

The user study included 63 healthy participants who are students and staff at the University of Texas at Arlington. The majority of participants are male (88.89%), right-handed (96.82%) participants. Table 5.2 summarizes the participants demographics. Each participant performed the baseline stage and 17 rounds of tasks (a round of T1, and 4-rounds of T2, T3, T4 and T5), excluding the trials. Also, they filled-out 18 surveys in total, a survey after the baseline stage and after every task round. Each experiment lasted between 1.5 to 2 hours.

[2] $https://github.com/abujelala/9PM$

Table 5.2: 9PM: Participants Demographics.

Participants Count	Ages	Gender	handedness
63	18-40 (Average =25.11, SD = 4.39)	Male : 56 Female : 7	Right-Handed: 61 Left-Handed: 2

5.6.2 Surveys Results and Analysis

This section discusses 9PM survey results. The survey questionnaires are available in Appendix A.1. These questionnaires were inspired by well-know survey questionnaires, such as: NASA Task Load Index (TLX) [44], Dundee Stress State Questionnaire [77], Chalder Fatigue Scale[49], and Positive and Negative Affect Schedule (PANAS) [122]. In addition, the participants provided feedback on the system's overall evaluation and the comfortability of the sensors. Their all overall evaluation was 8.7 out of 10 on average. Regarding the sensors, 44 participants found the EEG headset uncomfortable, 4 participants found the EDA electrodes uncomfortable, 3 participants found the IMU sensor uncomfortable, and only one participant found the ECG electrodes uncomfortable.

The 9PM surveys focus on 10 questions that ask about the participant's mental effort, physical effort, sleepiness/drowsiness, difficulty concentrating, task difficulty, task enjoyment, interest in the task, physical tiredness, distress, and attention. The statistical analysis of these survey questions is divided into two parts: the correlation between survey answers per task, and the variance between the responses across the tasks.

Table 5.3: 9PM: Survey Answers - Pearson's Correlation. The values represent Pearson's Linear Correlation Coefficient Rho ρ and *P value*. ρ (P value). Moderate positive correlations, with *P value* < 0.05, are marked in bold.

	Task 1	Task 2	Task 3	Task 4	Task 5
Mental Effort VS Physical Effort	**0.5021 (<0.0001)**	**0.564 (<0.0001)**	**0.5457 (<0.0001)**	**0.5628 (<0.0001)**	**0.6509 (<0.0001)**
Task Difficulty VS Mental Effort	**0.5204 (<0.0001)**	**0.4445 (0.0003)**	**0.3678 (0.003)**	**0.4511 (0.0002)**	0.2895 (0.0214)
Task Difficulty VS Physical Effort	**0.3617 (0.0036)**	0.2861 (0.023)	0.2421 (0.0559)	0.1829 (0.1512)	**0.4138 (0.0007)**
Physically Tired VS Physical Effort	0.2692 (0.0329)	0.2196 (0.0838)	0.1908 (0.1342)	0.2648 (0.036)	**0.3324 (0.0078)**
Feeling Attentive VS Physical Effort	0.0475 (0.7115)	0.2097 (0.0991)	0.0927 (0.4698)	0.2599 (0.0397)	0.1471 (0.2499)
Mental Effort VS Drowsy	-0.0437 (0.7337)	0.1776 (0.1638)	0.2011 (0.114)	-0.192 (0.1316)	-0.1129 (0.3783)
Task Enjoyment VS Mental Effort	0.2765 (0.0283)	0.2221 (0.0802)	0.1239 (0.3333)	0.1307 (0.3073)	0.2245 (0.0769)
Task Enjoyment VS Physical Effort	**0.3147 (0.012)**	0.2176 (0.0866)	0.2804 (0.026)	0.1633 (0.201)	0.1621 (0.2042)
Task Enjoyment VS Drowsy	-0.0745 (0.5615)	0.0499 (0.6976)	-0.1494 (0.2426)	-0.0739 (0.5646)	0.1333 (0.2978)
Task Enjoyment VS Task Difficulty	0.2478 (0.0503)	0.2056 (0.1059)	0.2055 (0.1061)	0.2867 (0.0227)	0.0041 (0.9748)

Correlation between Answers

There are 10 survey questions that we can investigate the correlations between their answers. That would lead to 45 comparison pairs. For simplicity, we focused on 10-pairs only. The correlations were checked using Pearson's correlation, Kendall Rank Correlation, and Spearman's Correlation. The three methods yielded similar results. Pearson's correlation results are available in Table 5.3, and the results of the other two methods are available in Appendix A.2.1. Generally, the results show that there is a moderate positive correlation (i.e., ρ or τ between 0.3 and 0.7 [100]) between mental effort and physical effort, and also between task difficulty and mental effort. In other words, when the participants felt that a task was mentally challenging, they also thought that they exerted more physical effort in the task. Thus, they marked the task as more difficult than tasks they found less mentally challenging.

Variance in Answers

To check if there is any variance in the answers across the tasks, we use the repeated measures One-Way Analysis Of Variance (ANOVA) test. The test was run on the answers of the 10 survey questions. Table 5.4 illustrates the summarized results. 'YES' means that there is significant variance in the answers, while 'No' refers to the lack of significant variance. The details of the ANOVA tests are available in Appendix A.2.1. In addition, this appendix also shows bar charts illustrating average survey responses, ± standard deviation (SD).

Table 5.4 shows that the significant variance in the answers is mainly limited to the questions asking about mental effort, task difficulty, and task enjoyment. Figures A.1 and A.2 show that the mental effort responses of T1 and T2 are similar, and they are slightly higher in T3. Also, T4 and T5 mental efforts responses are comparable, but they are significantly higher than the other tasks. Figures A.3 and A.4 show that the physical effort is very similar in almost all the tasks, with T1 requiring the highest physical effort. For task difficulty, T4 was reported the most difficult, followed by T5, T3, T2, and T1 (see Figure A.10). However, the difficulty of T1 and T2, and the difficulty of T4 and T5 are not significantly different (see Figure A.9). Task enjoyment responses were in ascending order, with T1 being the least enjoyed and T5 being the most enjoyed task (see Figure A.12). The details about the statistical analysis of variances are illustrated in Figure A.11.

Table 5.4: 9PM: Summary of the ANOVA Test Results on the Survey Answers. 'YES' indicates answers are significantly different while 'NO' indicate the lack of significant differences between the answers.

State	T1 vs T2	T1 vs T3	T1 vs T4	T1 vs T5	T2 vs T3	T2 vs T4	T2 vs T5	T3 vs T4	T3 vs T5	T4 vs T5
Mental Effort	NO	YES	YES	YES	NO	YES	YES	YES	YES	NO
Physical Effort	YES	YES	NO	NO	NO	NO	NO	NO	NO	NO
Sleepy or Drowsy	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
Difficulty Concentrating	NO	NO	YES	NO	NO	YES	YES	NO	NO	NO
Task Difficult	NO	YES	YES	YES	YES	YES	YES	YES	YES	NO
Task Enjoyment	NO	YES	YES	YES	YES	YES	YES	NO	NO	NO
Task Interesting	NO	NO	YES	YES	NO	YES	YES	NO	NO	NO
Feeling Physically Tired	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
Feeling Distressed	NO	NO	YES	NO	NO	YES	NO	YES	NO	NO
Feeling Attentive	NO	NO	NO	NO	NO	YES	YES	NO	NO	NO

5.6.3 Performance Results and Analysis

As mentioned in Section 5.3.3, performance data include task-score, move-time, and reaction-time. The participants were instructed to try to make as few mistakes as possible, and the task-scores reflect that. The average task-scores are 100%, 99.2%, 98.6%, 97.7% and 90.9% in T1, T2, T3, T4 and T5, respectively. In T4, the move after the rule changes randomly is not included in the task-score, since the participants would naturally make a mistake in this move. Also, incorrect moves that occurred due to external issues (e.g., pegs slipping from sweaty hands) are excluded from the performance analysis. This statistical analysis focuses on move-time and reaction time. Similar to Section 5.6.2, this analysis has two parts: the correlation between move-time and reaction-time, and the variance between the times.

Table 5.5: 9PM: Performance Analysis - Pearson's correlation, Kendall Rank Correlation, and Spearman's Correlation between the move-time and reaction-time in all the tasks. The values represent Correlation Coefficient Rho ρ (in case of Pearson and Spearman) or Tau τ (in case of Kendall) and the *P value*. ρ (P value) or τ (P value). Moderate positive correlations, with *P value* < 0.05, are marked in bold.

	T1	T2	T3	T4	T5
Pearson	**0.5824 (<0.0001)**	**0.5092 (<0.0001)**	**0.3591 (0.0038)**	**0.5938 (<0.0001)**	**0.555 (<0.0001)**
Kendall	**0.3917 (<0.0001)**	**0.3804 (<0.0001)**	0.2207 (<0.0108)	**0.3876 (<0.0001)**	**0.3743 (<0.0001)**
Spearman	**0.5486 (<0.0001)**	**0.5327 (<0.0001)**	**0.3402 (0.0066)**	**0.5483 (<0.0001)**	**0.5161 (<0.0001)**

Correlation between Move-Time and Reaction-Time

The correlations in section are also checked using Pearson's correlation, Kendall Rank Correlation, and Spearman's Correlation. Table 5.5 shows the correlations between the move-time and reaction-time. We have a moderate positive correlation in all the tasks, except in T3, it is a weak positive correlation when using Kendall Rank Correlation. The positive correlation might suggest that when reaction-time increases, the move-time also increases, and vice versa.

Table 5.6: 9PM: ANOVA Test on Move-Time and Reaction-Time. The values represent the *P value*, and significant differences (*P value* $<$ 0.05) are marked in bold.

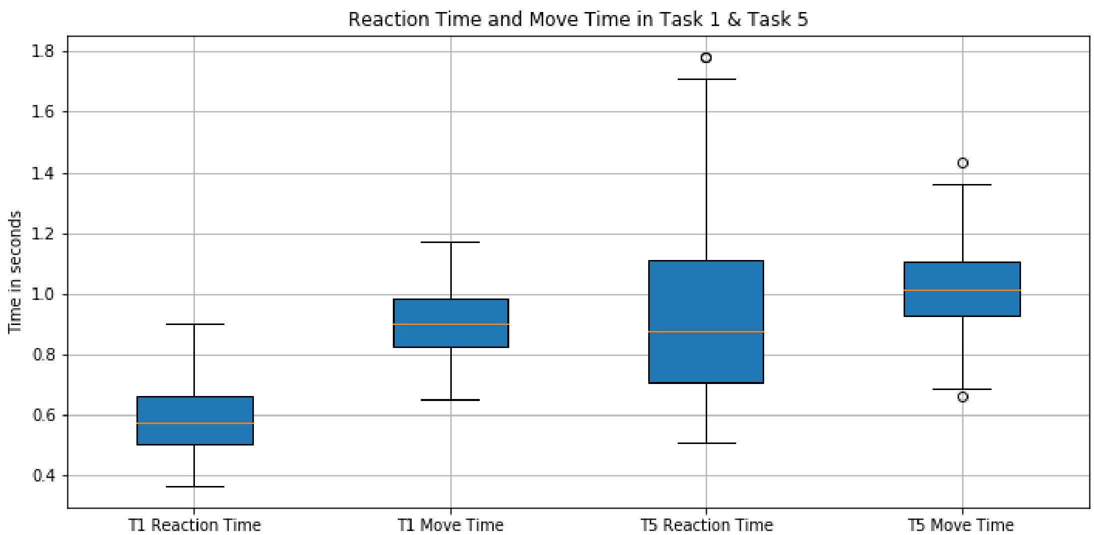
	Move-Time	**Reaction-Time**
ANOVA	<**0.0001**	<**0.0001**
T1 vs T2	0.0639	<**0.0001**
T1 vsT3	**0.0290**	<**0.0001**
T1 vsT4	<**0.0001**	<**0.0001**
T1 vsT5	<**0.0001**	<**0.0001**
T2 vs T3	0.6818	<**0.0001**
T2 vs T4	<**0.0001**	<**0.0001**
T2 vs T5	<**0.0001**	<**0.0001**
T3 vs T4	<**0.0001**	<**0.0001**
T3 vs T5	<**0.001**	<**0.0001**
T4 vs T5	0.1462	<**0.0001**

Variance in Time

To check if there is any variance in the times across the tasks, we use the repeated measures One-Way ANOVA test. The test is used twice: on the move-times, and the reaction-times. We also compared the reaction-time and move-time between T1 and T5 and between T2, T3, and T4. The tasks are divided into two groups because the reaction-time is calculated differently in these two groups, as explained in Section 5.3.3. Figures 5.7 and 5.8 show the comparison. The graphs show that when the reaction-time increases, the move-time also increases. Table 5.6 shows that there is a significant variance in the reaction-time across all the tasks. However, the move-time is not significantly different across all the tasks. In particular, the move-times between T1 and T2, T2 and T3, and T4 and

T5 are not significantly different. This finding is interesting since it matches the ANOVA test on mental effort responses; see Figure A.1. So, in the tasks where mental effort responses are not significantly different, the move-times are also not significantly different.

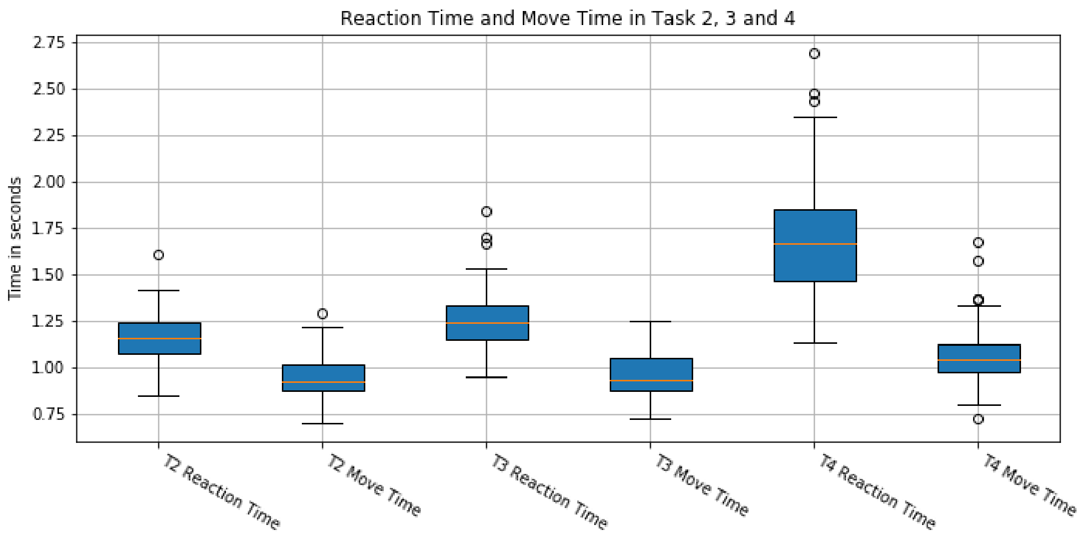
Table 5.7: 9PM: Boxplot showing the Reaction-Time and Move-Time in T1 and T5.



5.6.4 Machine Learning

We used Machine learning (ML) to utilize the user's multimodal data to predict the task the participant is performing and his/her survey answers. Our goal is to personalize the ML prediction based on the user's data, not based on the large bool of participants' data. Hence, each participant's data were normalized separately. We also converted the prediction of the survey answers into a classification problem, rather than a regression problem. In particular, we try to predict if the participant's response is higher than his/her average

Table 5.8: 9PM: Boxplot showing the Reaction-Time and Move-Time in T2, T3 and T4.



answers (e.g., exerted more than his/her average mental effort) or higher than the baseline response (e.g., feel more sleepy than in the baseline). There are 11 outcomes that ML tries to predict. Mental effort, physical effort, task difficulty, task enjoyment, task interesting are the 5-outcomes the ML tries to predict if the participant's response was higher than his/her average response. For example, the physical effort has a scale from 1 to 10. If the participant's lowest and highest responses are 3 and 9, respectively, the ML algorithm tries to predict if the participant's response in a particular task is higher than 6, based on his/her data. Sleepiness/drowsiness, difficulty concentrating, physical tiredness, distress, and attention are the 5-outcomes that we used ML to predict if they are higher than the baseline (e.g., the participant is feeling more attentive than at the beginning of the study). The final outcome to predict is the task number the participant is performing.

The multimodal data used to train the ML algorithms are a combination of

6 modalities. These modalities are ECG, EDA, EEG, IMU, performance metrics, and the task number, which lead to 63 possible combinations. When we try to predict the task number the participant is performing, the task number is not used as an input/feature in the ML. Each participant produced 17 datasets (one dataset per round), in addition to the baseline dataset. In T2, T3, T4, and T5, we averaged the rounds of each task to produced another 4 datasets. So, there are 21 datasets per participant presenting the tasks. All of these datasets are used to train, test, and validate the ML algorithms, as shown in Table 5.9. The data of 55 participants were used for training and validation using 5-folds cross-validation, and the data of the other 8 participants were used for testing.

Table 5.9: 9PM: Division of Participants' Datasets between ML Training, Validation and Testing.

63 Participants	
55 Participants (~87.3%)	8 Participants (~12.7%)
Training and Validation using Cross Validation (CV) with 5 Folds	Testing

We used 12 ML algorithms for classification, and we only report the results of the best classifier. These classifiers are tuned using the Randomized Grid-Search method provided by Scikit-Learn Library [105]. Randomized Grid-Search tests multiple variations of the algorithms (e.g., the same algorithm with various solvers and tuning methods), and it reports the effectiveness of each variation. Table 5.10 shows the 12 ML algorithms and their main variations.

To get the best results, we tried to two methods of feature normalization and two methods of feature selections. For normalization, we tried Min-Max Scaling and Standard Scaling. For Feature selection, one time, we used all the features, and the second time we used Principal Component Analysis (PCA) to select the

Table 5.10: 9PM: Machine Learning Classifiers.

Classifier	**Abbreviation**	**Main Variations**
Logistic Regression	LR	Solver: {lbfgs, liblinear, sag, saga}
K-Nearest Neighbors	KNN	Distance Metric: {euclidean, manhattan, minkowski}
Support Vector Machine	SVM	Kernel: {linear, poly, rbf, sigmoid}
Gradient Boosting	GB	
Extra Trees	ET	
Decision Tree	DT	Split Criterion: {gini, entropy}
Random Forest	RF	Split Criterion: {gini, entropy}
Neural Networks	NN	Solver: {sgd, adam, lbfgs} Activation: {identity, logistic, tanh, relu}
Naive Bayes	NB	
AdaBoost	AB	
Quadratic Discriminant Analysis	QDA	
Gaussian Process	GP	

features. For every method, we report the best classifier/algorithm, its F1-score on the test data, its accuracy on the test data, and average validation accuracy (average of the 5-folds). Since the labels are not equally distributed, we mainly focus on the F1-score while deciding the best results. Table 5.11 summaries the results of using the two scaling methods, with and without PCA. Also, it shows that F1-score on test data for Physical Effort, Sleepy/Drowsy, Difficulty Concentrating, and Physical Tiredness is below 70. This observation is expected since the ML analysis is based on the survey answers, and the answers of the participants for these questions were not significantly different, as shown in Table 5.4. In addition, we report the best ML results using single modalities (see Table 5.12) and combinations of modalities (see Appendix A.2.2).

Table 5.11: 9PM: Classification Results Using Two Scaling Methods with and without PCA. Best F1-scores are marked in bold.

	Min-Max Scaling				Standard Scaling			
	No PCA		PCA		No PCA		PCA	
Mental	AB	**77.50**	QDA	75.86	NB	75.45	NB	75.73
Effort	75.34	78.30	74.39	71.02	72.48	71.75	70.24	71.60
Physical	QDA	55.56	NB	43.84	QDA	**57.61**	NN	47.06
Effort	52.38	53.38	71.72	72.88	53.57	40.99	63.76	77.73
Sleepy	QDA	23.17	NB	26.67	QDA	26.42	NN	**28.57**
Drowsy	13.10	9.53	85.23	84.34	47.65	57.89	49.40	89.97
Difficulty	NB	48.98	QDA	**55.00**	QDA	49.54	QDA	48.65
Concentrating	70.24	63.89	75.34	69.67	66.67	60.77	74.50	67.68
Task	DT	85.71	AB	**86.67**	NB	82.35	NB	84.85
Difficulty	90.00	73.49	89.19	81.61	84.21	82.14	86.49	73.75
Task	LR	83.33	SVM	**84.21**	NN	83.33	KNN	81.08
Enjoyment	84.21	66.84	83.78	68.76	83.78	70.34	82.05	64.21
Task	KNN	73.33	NB	**76.92**	DT	72.73	GP	75.00
Interesting	78.95	67.34	77.50	52.75	76.32	61.28	74.36	52.17
Feeling	QDA	**50.00**	QDA	46.15	GP	44.44	QDA	47.06
Physically Tired	74.36	59.10	82.05	67.15	74.36	61.49	75.68	67.67
Feeling	GP	68.97	GP	66.67	GB	66.67	ET	**70.00**
Distressed	75.68	48.85	76.32	52.30	81.58	63.73	84.21	63.34
Feeling	GB	58.33	AB	**70.59**	QDA	57.14	RF	62.50
Attentive	74.36	62.04	86.84	61.00	75.68	50.30	84.62	57.78
Tasks	RF	79.88	NN	83.21	RF	80.62	LR	**83.37**
	80.41	70.64	83.56	69.00	81.08	73.50	83.56	71.28

In every Label vs. Dimensionality Reduction (e.g., Mental Effort vs. Min-Max Scaling No PCA):

– Upper-left cell = Best Classifier (e.g., AB) – Upper-Right cell = Test F1 Score (e.g., 77.50)

– Lower-left cell = Test Accuracy (e.g., 75.34) – Lower-Right cell = CV avg. Validation Accuracy(e.g., 78.30)

Table 5.12: 9PM: Classification Results from Single Modalities. Best F1-scores are marked in bold.

	ECG		EDA		EEG		IMU		Performance		Tasks	
Mental	SVM	68.53	DT	**76.43**	NB	64.89	NB	68.16	DT	65.56	NB	75.73
Effort	52.12	55.87	75.17	77.63	60.71	58.24	65.87	62.25	63.10	70.59	70.24	71.60
Physical	QDA	37.68	SVM	39.69	DT	**49.33**	NB	33.01	ET	25.35	AB	12.70
Effort	47.88	64.08	46.98	77.25	32.74	76.81	58.68	62.41	68.45	71.54	67.26	77.06
Sleepy	QDA	10.26	QDA	23.61	NN	**28.57**	GP	16.67	ET	6.06	QDA	18.69
Drowsy	78.79	81.69	26.17	52.12	49.40	89.97	76.05	79.89	81.55	86.77	48.21	78.43
Difficulty	NB	44.68	NB	38.10	GP	34.00	NB	41.43	ET	30.99	NB	**44.83**
Concentrating	68.48	62.16	56.38	58.29	60.71	53.62	50.90	50.18	70.83	68.14	61.90	60.35
Task	RF	64.52	ET	77.42	DT	76.92	NB	66.67	SVM	**81.25**	GP	75.00
Difficulty	71.79	58.90	81.58	86.66	77.50	69.08	67.50	69.35	85.00	77.83	80.00	90.55
Task	NN	**81.08**	DT	75.00	RF	74.42	RF	76.19	AB	61.90	ET	74.42
Enjoyment	82.05	63.30	78.95	60.43	72.50	60.41	75.00	62.05	60.00	64.40	72.50	73.51
Task	NB	62.22	DT	72.73	NB	**73.17**	NB	68.00	ET	62.50	NB	69.77
Interesting	56.41	52.00	76.32	61.28	72.50	52.77	60.00	50.37	70.00	56.38	67.50	63.32
Feeling	QDA	38.46	NN	34.78	GP	**44.44**	NB	36.36	GB	40.00	DT	33.33
Physically Tired	58.97	55.67	21.05	75.67	75.00	61.11	82.50	67.89	85.00	74.55	20.00	74.91
Feeling	NB	51.43	GB	57.14	GP	58.06	ET	**66.67**	LR	60.87	NB	55.56
Distressed	56.41	54.17	76.32	61.26	67.50	56.00	80.00	65.70	77.50	65.07	60.00	59.60
Feeling	RF	**62.50**	KNN	47.06	QDA	48.00	GB	53.85	AB	45.45	KNN	38.46
Attentive	84.62	57.78	76.32	61.24	67.50	56.42	70.00	63.11	70.00	62.53	60.00	62.18
Tasks	NN	45.66	NB	**77.01**	LR	69.04	ET	65.05	ET	68.10	N/A	N/A
	46.67	44.55	77.18	64.94	69.05	60.85	65.27	62.75	68.45	59.79	N/A	N/A

In every Label vs. Single Modality Box (e.g., Mental Effort vs. ECG):

– Upper-left cell = Best Classifier (e.g., SVM) – Upper-Right cell = Test F1 Score (e.g., 68.53)

– Lower-left cell = Test Accuracy (e.g., 52.12) – Lower-Right cell = CV avg. Validation Accuracy(e.g., 55.87)

5.7 Discussion and Conclusion

In this study, we were able to create the 9PM platform, a novel physical and cognitive assessment platform. It was designed based on the principles of standardized physical tests and cognitive tests. The platform collects multimodal data for analysis and assessment. We were able to analysis these multimodal data, and use them for assessment and answer our research questions. $Q1$ and $Q2$ ask if physiological, behavioral, and performance data be used in detection user's cognitive and physical performance, and which of these data is a better indicator. Table 5.11 shows that these data can be used to detect a user's cognitive and physical performance, and Table 5.12 shows their accuracy when the modalities are used separately. However, most of the time, these modalities give a more accurate indication when they are combined. Appendix A.2.2 has the results of the best 10 combinations. To answer $Q3$, we looked at how the reaction-time and move-time correlate. The reaction-time can be an indication of cognitive performance. We assume that a shorter reaction-time indicates better cognitive performance, and a shorter move-time indicates a better physical performance. Theoretically, since the participant makes the same physical action with disregard for the task, the move-time should not be significantly different from one task to another. However, that is not always the case, as shown in Table 5.6. Also, Table 5.5 shows that there is a moderate positive correlation between reaction-time and move-time. Base on that, we believe that there is a moderate positive correlation between cognitive performance and physical performance.

In our hypotheses, we expected that task difficulty would increase per task level (H_1) and that reaction-time would be significantly shorter in easier tasks

(H_2). We also anticipated that move-time would be similar in all the tasks (H_3) and that the collected data would be a good indicator of the participant's performance (H_4). H_1 did not hold fully true. Based on the survey responses (see Figure A.10), T4 was the most difficult task, followed by T5, T3, T2, and T1. ANOVA test (Figure A.9) shows that there was no significant difference in task difficulty between T1 and T2 and between T4 and T5. T1 did not include a cognitive task, and T2 had the easiest cognitive task. In T2, the participants had only to follow what is displayed on the screen without having to pay extra attention. Thus, the lack of cognitive element in T1 and the simplicity of the cognitive element in T2 might be the reason why the participants did not find a big difference in the difficulty between T1 and T2. From our previous work [119], we noticed that participants usually have difficulty with sequence memory tasks consisting of 9 items. Thus, we expected that T5 would be the most difficult task. However, in our previous work the sequence had 3 different options ('A', 'B', 'C'), but this study has only 2 options ('Blue', 'Red'). When the participants were asked on how they memorized the 9-items sequence, some of them mentioned that they only memorized the index of one color. For example, "Red Blue Blue Blue Red Red Blue Red Blue" would be memorized as "1 5 6 8" which represent the occurrence of 'Red' in the list. So the 9-items sequence is simplified into a 4-items sequence in this example. This can explain the reason why T5 was not the most difficult task as we expected.

H_2 hypothesis held true. The reaction-time was significantly different across the tasks (see Table 5.5) and it was shorter in easier tasks (see Figures 5.7, 5.8, A.10). H_3 did not hold fully true. Section 5.6.3 discussed how the move-time was not significantly different in tasks that required similar mental effort, but was significantly different otherwise. Finally, based on the results of Section

5.6.4 and the above-mentioned discussion on $Q3$, we find that H_4 did hold true; however, detecting physical workload/effort had lower accuracy due to the fact that the physical effort responses were not significantly different across most of the tasks.

CHAPTER 6

CONCLUDING REMARKS AND FUTURE DIRECTIONS

This dissertation explored a few HCI systems to monitor and assess human cognitive and physical abilities. These systems utilize robots, sensors, and wearable technologies.

In Chapter 1, we discussed the trends and challenges in human assessment, and we presented related work showing some of the commonly used physical and cognitive assessment tasks. Chapter 2 focused on robot-assisted assessment and rehabilitation. In particular, it presented two robot-assisted rehabilitation systems. The first system [20] uses a robotic arm for rehabilitation. The participants used the robotic arm to perform game-based rehabilitation exercises. This system also incorporates an eye-track to track user responses. Similarly, the second system [54] uses the same robotic arm for rehabilitation. However, this system analyzes EMG data from the participants to understand when they are physically fatigued. This system is adaptive, and the robotic arm behavior changes from resistive (challenging the participant) to assistive (helping the participant) when it predicts that the participant is fatigued. Chapter 3 discussed sleep-related studies briefly and focused on sleep-apnea. This chapter mainly discusses the results of our apnea detecting system [55]. This system is a prescreening tool for apnea in a home environment. It analyzes the data from an oximeter sensor to detect apnea events. Once apena events are detected, the system sends notifications with an advise to the user's smartwatch if the condition is severe. In Chapter 4, we explore brain-computer interfaces and show our system that uses EEG data to monitor user's engaged enjoyment. The users of this system play two games while they are wearing an EEG headset. The system then analyzes their EEG data to predict which game they like more. In Chapter 5, we discuss our novel multimodal cognitive and physical assessment platform, called ”9PM”. This system utilizes machine learning and multimodal data to assess user’s performance.

This dissertation is one step towards our goal to design intelligent systems that can monitor the user’s performance and provide personalized, adaptive feedback. One of the main steps to create such advanced systems is to understand the user’s current state. That is what our current systems illustrate in this dissertation. The next step would be make these system personalized and adaptive. That means the systems need to understand the user’s current state and provide feedback that would improve the user’s performance and life quality.

Finally, our systems provide successful proof of concepts. However, they still need to be validated and tuned using more substantial diverse datasets. The data used in this dissertation come from healthy participants in a controlled lab environment. Thus, our systems still need to be tested and validated on data that represent real-life scenarios.

APPENDIX A

APPENDIX: 9PM STUDY MATERIALS

A.1 User Survey

A.1.1 9PM: Baseline Survey

1. Enter your age:

   [e.g., 27]

2. Enter your gender:

   [e.g., Male]

3. Are you right-handed or left-handed? Select one:

   (a) Right-Handed

   (b) Left-Handed

4. Do you feel sleepy or drowsy at the moment? Select one:

   (a) Less than Usual

   (b) No More than Usual

   (c) More than Usual

   (d) Much More than Usual

5. Do you have difficulties concentrating at the moment? Select one:

   (a) Less than Usual

   (b) No More than Usual

(c) More than Usual
(d) Much More than Usual

6. Do you feel physically tired at the moment? Select one:
   (a) Less than Usual
   (b) No More than Usual
   (c) More than Usual
   (d) Much More than Usual
7. How distressed do you feel at the moment? Select one:
   (a) Very Slightly or Not at All
   (b) A Little
   (c) Moderately
   (d) Quite a Bit
   (e) Extremely
8. How attentive do you feel at the moment? Select one:
   (a) Very Slightly or Not at All
   (b) A Little
   (c) Moderately
   (d) Quite a Bit
   (e) Extremely

A.1.2 9PM: Round Survey

1. How much mental effort did you spend on this task? Select a number from 1 (*Very Low*) to 10 (*Very High*):
   [e.g., 7]
2. How much physical effort did you spend on this task? Select a number from 1 (*Very Low*) to 10 (*Very High*):
   [e.g., 7]
3. Did you feel sleepy or drowsy during the task? Select one:
   (a) Less than Usual
   (b) No More than Usual
   (c) More than Usual
   (d) Much More than Usual
4. Did you have difficulties concentrating during the task?? Select one:
   (a) Less than Usual
   (b) No More than Usual
   (c) More than Usual
   (d) Much More than Usual

A.1.3 9PM: Task Survey

1. How much mental effort did you spend on this task? Select a number from 1 (*Very Low*) to 10 (*Very High*):
   [e.g., 7]
2. How much physical effort did you spend on this task? Select a number from 1 (*Very Low*) to 10 (*Very High*):
   [e.g., 7]
3. Did you feel sleepy or drowsy during the task? Select one:
   (a) Less than Usual
   (b) No More than Usual
   (c) More than Usual
   (d) Much More than Usual
4. Did you have difficulties concentrating during the task?? Select one:
   (a) Less than Usual
   (b) No More than Usual
   (c) More than Usual
   (d) Much More than Usual
5. Did you feel physically tired during the task? Select one:
   (a) Less than Usual
   (b) No More than Usual
   (c) More than Usual
   (d) Much More than Usual

6. How distressed did you feel during the task?? Select one:
   (a) Very Slightly or Not at All
   (b) A Little
   (c) Moderately
   (d) Quite a Bit
   (e) Extremely
7. How interested were you in the task? Select one:
   (a) Very Slightly or Not at All
   (b) A Little
   (c) Moderately
   (d) Quite a Bit
   (e) Extremely
8. How attentive were you during the task? Select one:
   (a) Very Slightly or Not at All
   (b) A Little
   (c) Moderately
   (d) Quite a Bit
   (e) Extremely

A.1.4 9PM: Experiment Feedback Survey

1. How difficult was Task 1? Select a number from 1 (*Very Easy*) to 10 (*Very Difficult*):

   [e.g., 7]

2. How difficult was Task 2? Select a number from 1 (*Very Easy*) to 10 (*Very Difficult*):

   [e.g., 7]

3. How difficult was Task 3? Select a number from 1 (*Very Easy*) to 10 (*Very Difficult*):

   [e.g., 7]

4. How difficult was Task 4? Select a number from 1 (*Very Easy*) to 10 (*Very Difficult*):

   [e.g., 7]

5. How difficult was Task 5? Select a number from 1 (*Very Easy*) to 10 (*Very Difficult*):

   [e.g., 7]

6. How much did you enjoy Task 1? Select a number from 1 (*Not at All*) to 10 (*Very Much*):

   [e.g., 7]

7. How much did you enjoy Task 2? Select a number from 1 (*Not at All*) to 10 (*Very Much*):

   [e.g., 7]

8. How much did you enjoy Task 3? Select a number from 1 (*Not at All*) to 10 (*Very Much*):

   [e.g., 7]

9. How much did you enjoy Task 4? Select a number from 1 (*Not at All*) to 10 (*Very Much*):

   [e.g., 7]

10. How much did you enjoy Task 5? Select a number from 1 (*Not at All*) to 10 (*Very Much*):

    [e.g., 7]

11. Which sensors were uncomfortable to wear? Please check all that apply. You may check more than one option.

    (a) EEG Headset

    (b) ECG Sensor

    (c) EDA Sensor

    (d) IMU Sensor

    (e) None of the above

12. What is your overall evaluation of the experiment? Select a number from 1 (*Very Bad*) to 10 (*Very Good*):

    [e.g., 7]

13. Please write any other comments you may have:

A.2 Results and Analysis

A.2.1 Survey Results

Correlation Analysis on Survey Answers

Table A.1: 9PM: Survey Answers - Kendall Correlation. The values represent Kendall's Tau Coefficient τ and $P - Value$. τ (P-Value). Strong and moderate positive correlations, with $P{-}Value < 0.05$, are marked in bold letters.

	Task 1	Task 2	Task 3	Task 4	Task 5
Mental Effort VS Physical Effort	**0.4183 (<0.0001)**	**0.4636 (<0.0001)**	**0.4528 (<0.0001)**	**0.4129 (<0.0001)**	**0.4939 (<0.0001)**
Task Difficulty VS Mental Effort	**0.4343 (<0.0001)**	**0.3933 (<0.0001)**	**0.3156 (0.0008)**	**0.3521 (0.0002)**	0.2408 (0.0093)
Task Difficulty VS Physical Effort	0.2518 (0.0141)	0.2447 (0.0123)	0.1897 (0.0458)	0.1368 (0.1549)	0.2918 (0.0019)
Physically Tired VS Physical Effort	0.1598 (0.1247)	0.227 (0.0278)	0.2176 (0.0337)	0.1894 (0.0633)	0.2157 (0.0348)
Feeling Attentive VS Physical Effort	0.089 (0.3928)	0.1297 (0.1976)	0.0505 (0.6125)	0.1821 (0.0684)	0.1015 (0.3018)
Mental Effort VS Drowsy	-0.0253 (0.8147)	0.1281 (0.1903)	0.1532 (0.1139)	-0.1234 (0.2029)	-0.0732 (0.4419)
Task Enjoyment VS Mental Effort	0.2078 (0.0315)	0.1588 (0.0902)	0.0503 (0.5954)	0.1136 (0.2236)	0.1786 (0.0548)
Task Enjoyment VS Physical Effort	0.1923 (0.0472)	0.1762 (0.0638)	0.1766 (0.0612)	0.0845 (0.3713)	0.0909 (0.3368)
Task Enjoyment VS Drowsy	-0.0746 (0.4782)	0.0594 (0.5564)	-0.1017 (0.3096)	-0.02 (0.8469)	0.0981 (0.3223)
Task Enjoyment VS Task Difficulty	0.2068 (0.0421)	0.1669 (0.0919)	0.1351 (0.1664)	0.2922 (0.0031)	0.0182 (0.8553)

Table A.2: 9PM: Survey Answers - Spearman Correlation. The values represent Spearman's Rho Coefficient ρ and $P - Value$. ρ (P-Value). Strong and moderate positive correlations, with $P{-}Value < 0.05$, are marked in bold letters.

	Task 1	Task 2	Task 3	Task 4	Task 5
Mental Effort VS Physical Effort	**0.5392** **(<0.0001)**	**0.5909** **(<0.0001)**	**0.5736** **(<0.0001)**	**0.5192** **(<0.0001)**	**0.6489** **(<0.0001)**
Task Difficulty VS Mental Effort	**0.5282** **(<0.0001)**	**0.504** **(<0.0001)**	**0.412** **(0.0008)**	**0.4471** **(0.0002)**	**0.3129** **(0.0125)**
Task Difficulty VS Physical Effort	**0.3022** **(0.0161)**	**0.3185** **(0.0109)**	0.246 (0.052)	0.165 (0.1963)	**0.3782** **(0.0022)**
Physically Tired VS Physical Effort	0.1976 (0.1205)	0.2793 (0.0266)	0.2535 (0.045)	0.2269 (0.0737)	0.2584 (0.0409)
Feeling Attentive VS Physical Effort	0.112 (0.3823)	0.1639 (0.1993)	0.0733 (0.5679)	0.2317 (0.0677)	0.1287 (0.3148)
Mental Effort VS Drowsy	-0.0297 (0.8173)	0.1617 (0.2054)	0.1895 (0.1368)	-0.1687 (0.1862)	-0.1006 (0.4326)
Task Enjoyment VS Mental Effort	0.2741 (0.0297)	0.2449 (0.0531)	0.0706 (0.5825)	0.1506 (0.2386)	0.2411 (0.057)
Task Enjoyment VS Physical Effort	0.2551 (0.0436)	0.2304 (0.0692)	0.2371 (0.0614)	0.1118 (0.383)	0.1264 (0.3234)
Task Enjoyment VS Drowsy	-0.0943 (0.4624)	0.064 (0.6184)	-0.1125 (0.3801)	-0.0168 (0.8958)	0.128 (0.3176)
Task Enjoyment VS Task Difficulty	0.2678 (0.0338)	0.2113 (0.0965)	0.1677 (0.1888)	**0.3605** **(0.0037)**	0.0009 (0.9943)

Significant Difference Analysis on Survey Answers

Mental Effort

Anova

	F Value	Num DF	Den DF	Pr > F
Round	70.1449	4.0000	248.0000	0.0000

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
T1/R1	T2	0.127	0.9	-0.8815	1.1355	False
T1/R1	T3	1.1151	0.0219	0.1066	2.1236	True
T1/R1	T4	2.6032	0.001	1.5947	3.6117	True
T1/R1	T5	2.6111	0.001	1.6026	3.6196	True
T2	T3	0.9881	0.058	-0.0204	1.9966	False
T2	T4	2.4762	0.001	1.4677	3.4847	True
T2	T5	2.4841	0.001	1.4756	3.4926	True
T3	T4	1.4881	0.001	0.4796	2.4966	True
T3	T5	1.496	0.001	0.4875	2.5045	True
T4	T5	0.0079	0.9	-1.0006	1.0164	False

Figure A.1: 9PM: Survey Answers - ANOVA Test on Mental Effort Responses.

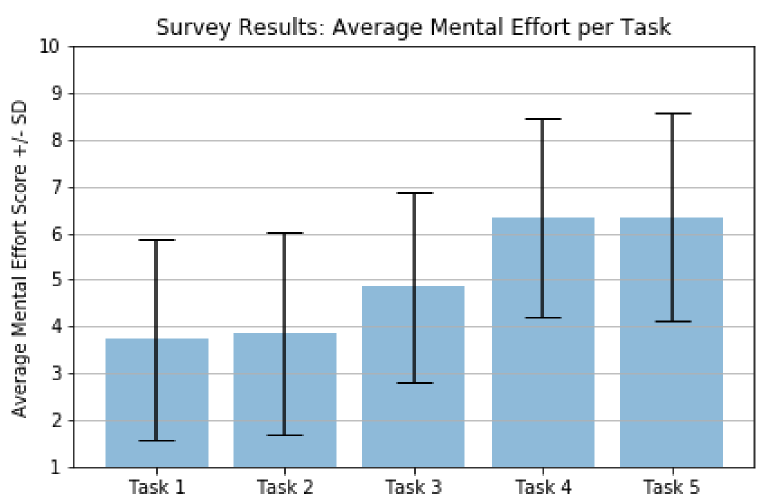


Figure A.2: 9PM: Survey Answers - Average Mental Effort Responses.

Physical Effort

Anova

	F Value	Num DF	Den DF	Pr > F
Round	14.0394	4.0000	248.0000	0.0000

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
T1/R1	T2	-1.2778	0.0013	-2.1858	-0.3697	True
T1/R1	T3	-1.004	0.0219	-1.912	-0.0959	True
T1/R1	T4	-0.6706	0.2557	-1.5787	0.2374	False
T1/R1	T5	-0.6071	0.356	-1.5152	0.3009	False
T2	T3	0.2738	0.9	-0.6342	1.1818	False
T2	T4	0.6071	0.356	-0.3009	1.5152	False
T2	T5	0.6706	0.2557	-0.2374	1.5787	False
T3	T4	0.3333	0.8335	-0.5747	1.2414	False
T3	T5	0.3968	0.725	-0.5112	1.3049	False
T4	T5	0.0635	0.9	-0.8445	0.9715	False

Figure A.3: 9PM: Survey Answers - ANOVA Test on Physical Effort Responses.

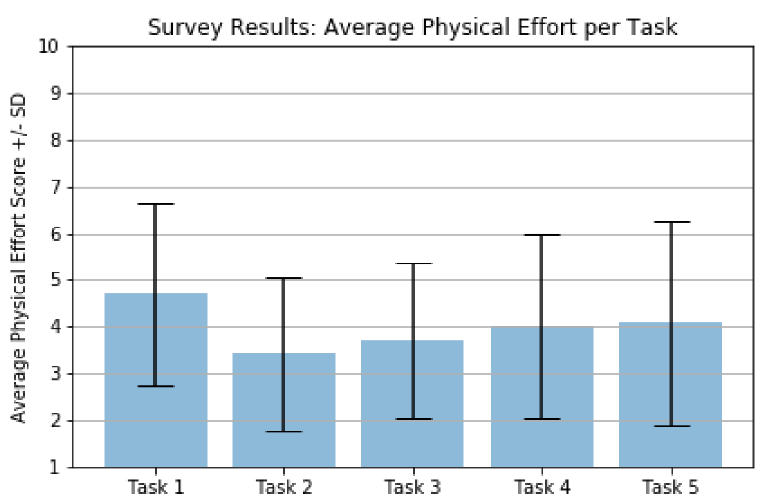


Figure A.4: 9PM: Survey Answers - Average Physical Effort Responses.

Sleepy / Drowsy

Anova

	F Value	Num DF	Den DF	Pr > F
Round	0.8023	4.0000	248.0000	0.5247

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
T1/R1	T2	0.0794	0.9	-0.2889	0.4476	False
T1/R1	T3	0.0595	0.9	-0.3087	0.4278	False
T1/R1	T4	0.0437	0.9	-0.3246	0.4119	False
T1/R1	T5	0.0873	0.9	-0.281	0.4556	False
T2	T3	-0.0198	0.9	-0.3881	0.3484	False
T2	T4	-0.0357	0.9	-0.404	0.3326	False
T2	T5	0.0079	0.9	-0.3603	0.3762	False
T3	T4	-0.0159	0.9	-0.3841	0.3524	False
T3	T5	0.0278	0.9	-0.3405	0.396	False
T4	T5	0.0437	0.9	-0.3246	0.4119	False

Figure A.5: 9PM: Survey Answers - ANOVA Test on Sleepy/Drowsy Responses.

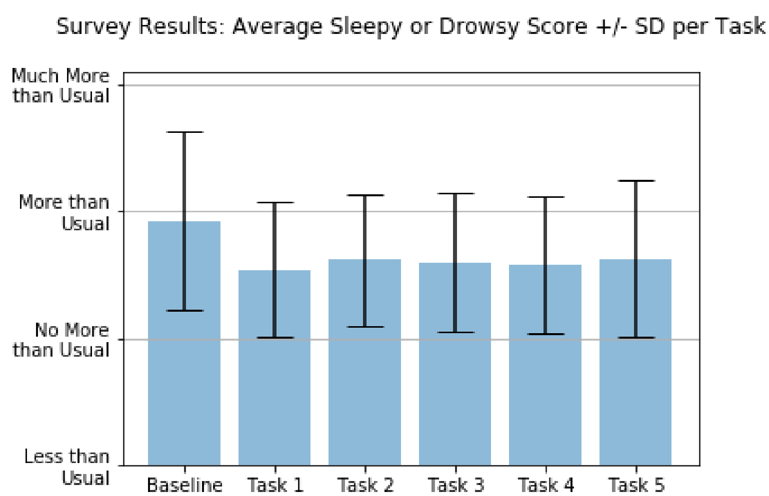


Figure A.6: 9PM: Survey Answers - Average Sleepy/Drowsy Responses.

Difficulty Concentrating

Anova

	F Value	Num DF	Den DF	Pr > F
Round	13.6025	4.0000	248.0000	0.0000

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
T1/R1	T2	-0.0079	0.9	-0.3522	0.3363	False
T1/R1	T3	0.0714	0.9	-0.2728	0.4157	False
T1/R1	T4	0.4087	0.0109	0.0645	0.753	True
T1/R1	T5	0.3373	0.058	-0.007	0.6816	False
T2	T3	0.0794	0.9	-0.2649	0.4236	False
T2	T4	0.4167	0.0088	0.0724	0.7609	True
T2	T5	0.3452	0.049	0.001	0.6895	True
T3	T4	0.3373	0.058	-0.007	0.6816	False
T3	T5	0.2659	0.2146	-0.0784	0.6101	False
T4	T5	-0.0714	0.9	-0.4157	0.2728	False

Figure A.7: 9PM: Survey Answers - ANOVA Test on Difficulty Concentrating Responses.

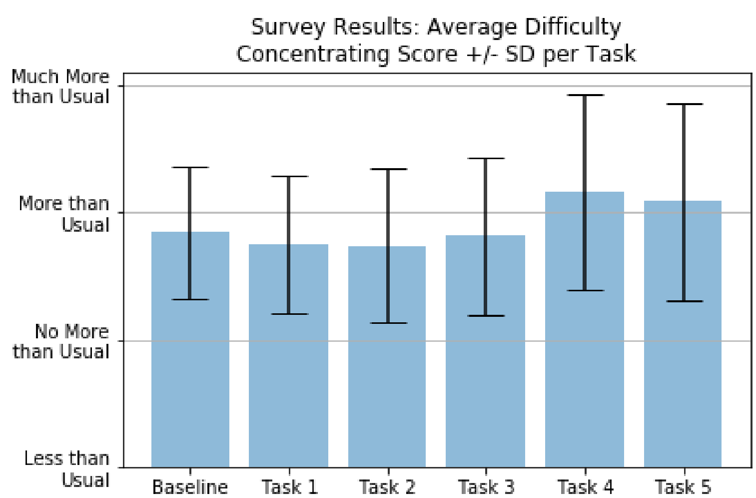


Figure A.8: 9PM: Survey Answers - Average Difficulty Concentrating Responses.

Task Difficulty

```
                  Anova
=====================================
      F Value  Num DF  Den DF  Pr > F
-------------------------------------
Round 131.1968 4.0000 248.0000 0.0000
=====================================

Multiple Comparison of Means - Tukey HSD, FWER=0.05
===================================================
group1 group2 meandiff p-adj   lower  upper  reject
---------------------------------------------------
 T1/R1     T2   0.6032  0.368 -0.3092 1.5155  False
 T1/R1     T3   1.6984  0.001  0.7861 2.6107   True
 T1/R1     T4   4.9206  0.001  4.0083  5.833   True
 T1/R1     T5   4.1905  0.001  3.2781 5.1028   True
    T2     T3   1.0952 0.0097  0.1829 2.0076   True
    T2     T4   4.3175  0.001  3.4051 5.2298   True
    T2     T5   3.5873  0.001   2.675 4.4996   True
    T3     T4   3.2222  0.001  2.3099 4.1346   True
    T3     T5   2.4921  0.001  1.5797 3.4044   True
    T4     T5  -0.7302 0.1838 -1.6425 0.1822  False
---------------------------------------------------
```

Figure A.9: 9PM: Survey Answers - ANOVA Test on Task Difficulty Responses.

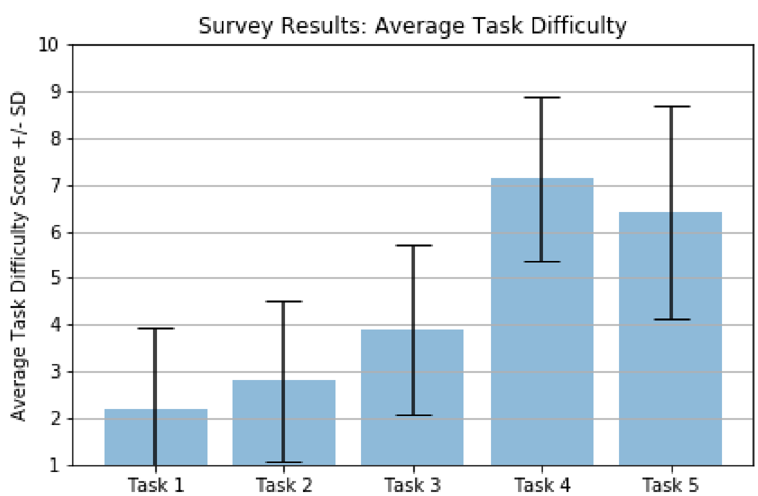


Figure A.10: 9PM: Survey Answers - Average Task Difficulty Responses.

Task Enjoyment

```
                 Anova
====================================
      F Value Num DF  Den DF  Pr > F
------------------------------------
Round 29.9354 4.0000 248.0000 0.0000
====================================

Multiple Comparison of Means - Tukey HSD, FWER=0.05
===================================================
group1 group2 meandiff p-adj   lower  upper  reject
---------------------------------------------------
 T1/R1     T2   0.8571 0.2554  -0.303 2.0173  False
 T1/R1     T3    2.381  0.001  1.2208 3.5411   True
 T1/R1     T4   2.8889  0.001  1.7287  4.049   True
 T1/R1     T5   2.9841  0.001   1.824 4.1443   True
    T2     T3   1.5238 0.0033  0.3637  2.684   True
    T2     T4   2.0317  0.001  0.8716 3.1919   True
    T2     T5    2.127  0.001  0.9668 3.2871   True
    T3     T4   0.5079 0.7237 -0.6522 1.6681  False
    T3     T5   0.6032 0.5963  -0.557 1.7633  False
    T4     T5   0.0952    0.9 -1.0649 1.2554  False
---------------------------------------------------
```

Figure A.11: 9PM: Survey Answers - ANOVA Test on Task Enjoyment Responses.

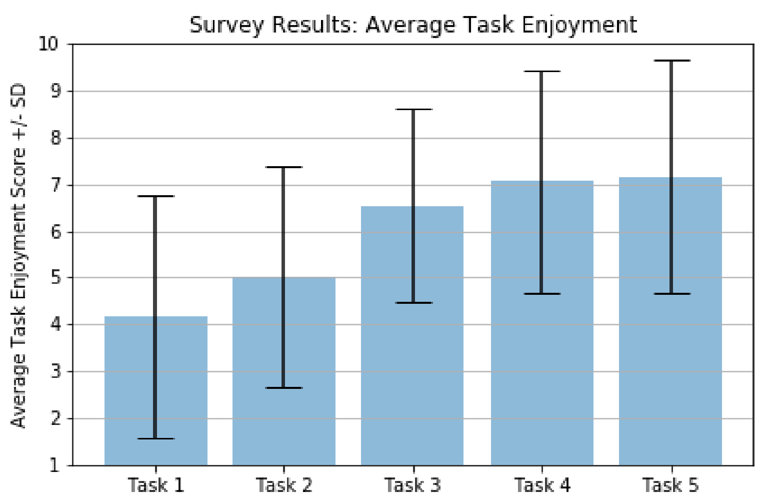


Figure A.12: 9PM: Survey Answers - Average Task Enjoyment Responses.

Task Interesting

```
                  Anova
====================================
      F Value Num DF  Den DF  Pr > F
------------------------------------
Round 10.9677 4.0000 248.0000 0.0000
====================================
```

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
T1/R1	T2	-0.0794	0.9	-0.5921	0.4334	False
T1/R1	T3	0.3016	0.4897	-0.2111	0.8143	False
T1/R1	T4	0.619	0.0091	0.1063	1.1318	True
T1/R1	T5	0.5397	0.0335	0.027	1.0524	True
T2	T3	0.381	0.25	-0.1318	0.8937	False
T2	T4	0.6984	0.0021	0.1857	1.2111	True
T2	T5	0.619	0.0091	0.1063	1.1318	True
T3	T4	0.3175	0.4378	-0.1953	0.8302	False
T3	T5	0.2381	0.6825	-0.2746	0.7508	False
T4	T5	-0.0794	0.9	-0.5921	0.4334	False

Figure A.13: 9PM: Survey Answers - ANOVA Test on Task Interesting Responses.

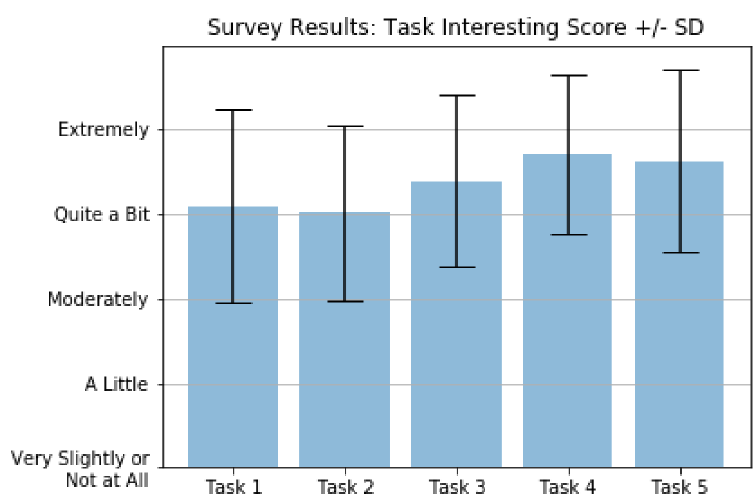


Figure A.14: 9PM: Survey Answers - Average Task Interesting Responses.

Physical Tiredness

```
                 Anova
====================================
      F Value Num DF  Den DF  Pr > F
------------------------------------
Round  1.1002 4.0000 248.0000 0.3570
====================================

Multiple Comparison of Means - Tukey HSD, FWER=0.05
===================================================
group1 group2 meandiff p-adj   lower  upper  reject
---------------------------------------------------
 T1/R1     T2      0.0    0.9 -0.3805 0.3805  False
 T1/R1     T3   0.0794    0.9 -0.3012 0.4599  False
 T1/R1     T4   0.1111    0.9 -0.2694 0.4916  False
 T1/R1     T5    0.127 0.8853 -0.2536 0.5075  False
    T2     T3   0.0794    0.9 -0.3012 0.4599  False
    T2     T4   0.1111    0.9 -0.2694 0.4916  False
    T2     T5    0.127 0.8853 -0.2536 0.5075  False
    T3     T4   0.0317    0.9 -0.3488 0.4123  False
    T3     T5   0.0476    0.9 -0.3329 0.4282  False
    T4     T5   0.0159    0.9 -0.3647 0.3964  False
---------------------------------------------------
```

Figure A.15: 9PM: Survey Answers - ANOVA Test on Physical Tiredness Responses.

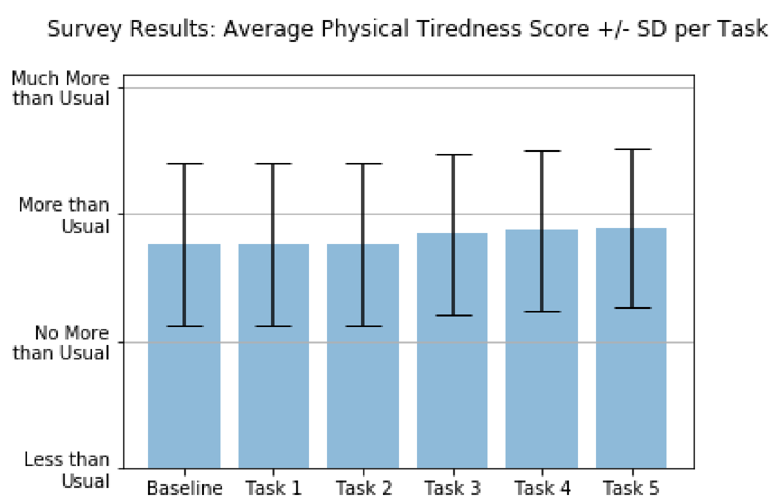


Figure A.16: 9PM: Survey Answers - Average Physical Tiredness Responses.

Distress

```
                Anova
====================================
      F Value Num DF  Den DF  Pr > F
------------------------------------
Round 17.6969 4.0000 248.0000 0.0000
====================================

Multiple Comparison of Means - Tukey HSD, FWER=0.05
===================================================
group1 group2 meandiff p-adj   lower  upper  reject
---------------------------------------------------
 T1/R1     T2      0.0    0.9 -0.5327 0.5327  False
 T1/R1     T3    0.254 0.6633 -0.2788 0.7867  False
 T1/R1     T4   0.8571  0.001  0.3244 1.3899   True
 T1/R1     T5   0.4921 0.0859 -0.0407 1.0248  False
    T2     T3    0.254 0.6633 -0.2788 0.7867  False
    T2     T4   0.8571  0.001  0.3244 1.3899   True
    T2     T5   0.4921 0.0859 -0.0407 1.0248  False
    T3     T4   0.6032 0.0175  0.0705 1.1359   True
    T3     T5   0.2381 0.7096 -0.2946 0.7708  False
    T4     T5  -0.3651   0.33 -0.8978 0.1676  False
---------------------------------------------------
```

Figure A.17: 9PM: Survey Answers - ANOVA Test on Feeling Distressed Responses.

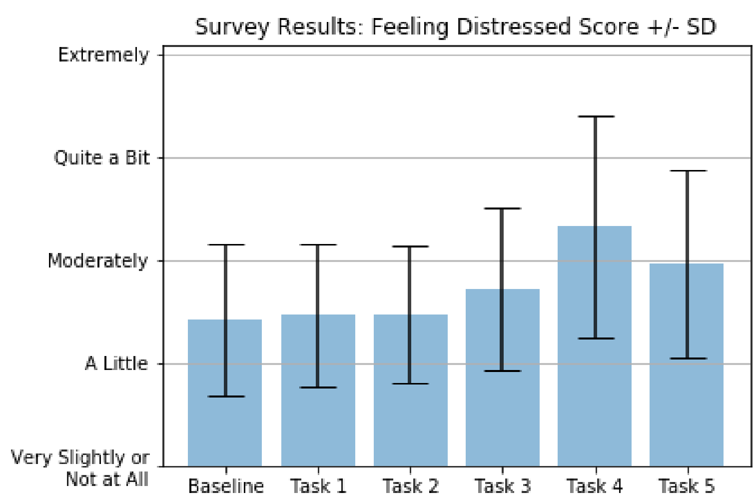


Figure A.18: 9PM: Survey Answers - Average Distress Responses.

Attention

```
                    Anova
====================================
      F Value Num DF  Den DF   Pr > F
------------------------------------
Round 11.5906 4.0000 248.0000 0.0000
====================================

Multiple Comparison of Means - Tukey HSD, FWER=0.05
===================================================
group1 group2 meandiff p-adj   lower  upper  reject
---------------------------------------------------
 T1/R1     T2   -0.254 0.6563 -0.7817 0.2737  False
 T1/R1     T3  -0.1746 0.8897 -0.7023 0.3531  False
 T1/R1     T4   0.3333 0.4167 -0.1944  0.861  False
 T1/R1     T5   0.3175 0.4673 -0.2102 0.8451  False
    T2     T3   0.0794    0.9 -0.4483 0.6071  False
    T2     T4   0.5873 0.0206  0.0596  1.115   True
    T2     T5   0.5714 0.0263  0.0437 1.0991   True
    T3     T4   0.5079 0.0657 -0.0198 1.0356  False
    T3     T5   0.4921  0.081 -0.0356 1.0198  False
    T4     T5  -0.0159    0.9 -0.5436 0.5118  False
---------------------------------------------------
```

Figure A.19: 9PM: Survey Answers - ANOVA Test on Attention Responses.

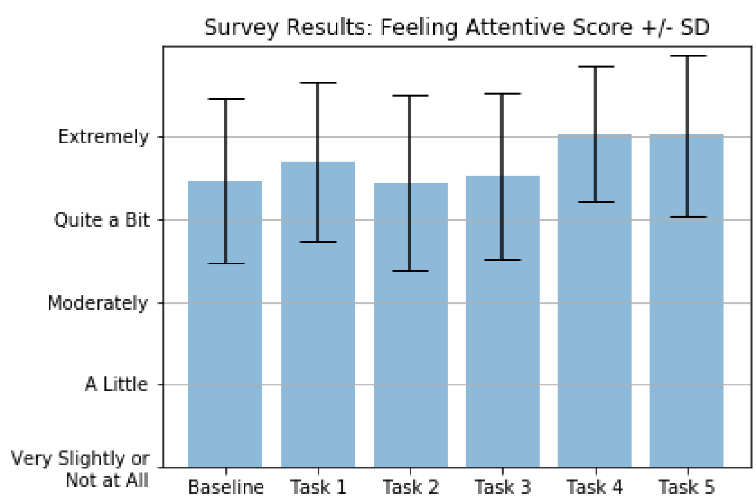


Figure A.20: 9PM: Survey Answers - Average Attention Responses.

A.2.2 Machine Learning Results

Table A.3: 9PM: 10 Best Modalities for Mental Effort Classification.

Best 10 Modalities	Best Classifier	Test F1	Test Accuracy	CV Avg. Validation Accuracy ± SD
ECG+EDA+Performance	AB	77.50	75.34	78.30 ± 3.92
EEG+EDA+Performance	RF	76.92	75.84	77.97 ± 5.57
EDA+Performance	AB	76.92	75.84	78.77 ± 4.92
EEG+EDA	GB	76.62	75.84	75.50 ± 4.75
EEG+ECG+EDA+IMU+Performance	RF	76.62	75.17	78.17 ± 5.43
EEG+EDA+IMU+Performance	RF	76.62	75.68	77.79 ± 4.97
EEG+ECG+EDA+Performance	GB	76.62	75.34	76.83 ± 4.62
EDA	DT	76.43	75.17	77.63 ± 5.60
EEG+ECG+EDA	DT	76.43	74.66	77.59 ± 5.89
ECG+EDA	DT	76.43	74.66	77.63 ± 5.60

Table A.4: 9PM: 10 Best Modalities for Physical Effort Classification.

Best 10 Modalities	Best Classifier	Test F1	Test Accuracy	CV Avg. Validation Accuracy ± SD
Performance+Tasks	QDA	57.61	53.57	40.99 ± 17.73
ECG+Tasks	QDA	50.00	61.21	68.61 ± 3.84
EEG+IMU+Tasks	QDA	49.76	37.13	22.68 ± 0.59
EEG+ECG+IMU+Performance+Tasks	QDA	49.47	41.46	26.39 ± 7.01
EEG+Tasks	DT	49.33	32.74	76.81 ± 0.09
EEG+Performance	DT	49.33	32.74	76.81 ± 0.09
EEG+Performance+Tasks	DT	49.33	32.74	76.81 ± 0.09
EEG	DT	49.33	32.74	76.81 ± 0.09
EEG+ECG+IMU+Performance	QDA	48.85	32.32	29.70 ± 10.58
EEG+ECG+IMU+Tasks	QDA	48.85	32.32	28.48 ± 8.64

Table A.5: 9PM: 10 Best Modalities for Sleepy/Drowsy Classification.

Best 10 Modalities	Best Classifier	Test F1	Test Accuracy	CV Avg. Validation Accuracy ± SD
EEG	NN	28.57	49.40	89.97 ± 0.02
EEG+EDA+Tasks	NB	26.67	85.23	84.34 ± 3.39
EEG+EDA+Performance+Tasks	NB	26.67	77.85	79.49 ± 5.26
EEG+ECG+EDA	NB	26.67	77.40	79.59 ± 4.90
EEG+EDA+Performance	NB	26.67	77.85	79.68 ± 5.29
EEG+ECG+EDA+Tasks	NB	26.67	77.40	79.30 ± 4.97
EDA+Performance	QDA	26.42	47.65	57.89 ± 12.86
EEG+EDA	NB	25.00	75.84	79.40 ± 5.24
EEG+ECG+IMU+Performance+Tasks	NB	24.62	70.12	77.81 ± 2.56
EEG+ECG+IMU+Performance	NB	24.62	70.12	77.98 ± 2.50

Table A.6: 9PM: 10 Best Modalities for Concentrating Difficulty Classification.

Best 10 Modalities	Best Classifier	Test F1	Test Accuracy	CV Avg. Validation Accuracy ± SD
ECG+EDA+Tasks	QDA	55.00	75.34	69.67 ± 2.66
ECG+EDA+IMU+Performance+Tasks	NB	54.55	79.31	64.71 ± 3.33
EDA+Tasks	QDA	54.55	76.51	68.63 ± 2.93
EDA+Performance+Tasks	QDA	54.32	75.17	69.76 ± 3.08
ECG+EDA+Performance+Tasks	QDA	54.29	78.08	71.09 ± 1.50
ECG+EDA+IMU	QDA	53.16	74.48	59.56 ± 6.12
ECG+EDA+IMU+Performance	QDA	51.95	74.48	59.66 ± 6.88
ECG+EDA+IMU+Tasks	LR	50.91	81.38	66.70 ± 2.66
EDA+IMU	QDA	50.63	73.65	59.76 ± 6.24
ECG+Performance+Tasks	QDA	49.54	66.67	60.77 ± 2.25

Table A.7: 9PM: 10 Best Modalities for Task Difficulty Classification.

Best 10 Modalities	Best Classifier	Test F1	Test Accuracy	CV Avg. Validation Accuracy ± SD
ECG+EDA+IMU+Tasks	AB	86.67	89.19	81.61 ± 6.49
EEG+Performance	DT	85.71	90.00	73.49 ± 5.91
ECG+EDA+Performance+Tasks	NB	84.85	86.49	73.75 ± 4.99
ECG+EDA+IMU+Performance+Tasks	GB	83.87	86.49	86.20 ± 6.97
ECG+IMU+Tasks	QDA	83.87	87.18	67.88 ± 5.24
EDA+IMU+Tasks	AB	83.87	86.84	83.58 ± 1.64
ECG+IMU+Performance+Tasks	QDA	83.87	87.18	68.98 ± 5.40
ECG+Performance	DT	83.87	87.18	78.56 ± 6.42
EEG+ECG+EDA+Tasks	QDA	82.76	86.49	65.78 ± 2.04
ECG+EDA+Tasks	GB	82.76	86.49	87.07 ± 4.19

Table A.8: 9PM: 10 Best Modalities for Task Enjoyment Classification.

Best 10 Modalities	Best Classifier	Test F1	Test Accuracy	CV Avg. Validation Accuracy ± SD
ECG+EDA+IMU+Performance+Tasks	SVM	84.21	83.78	68.76 ± 7.84
EDA+IMU+Performance+Tasks	LR	83.33	84.21	66.84 ± 5.55
EEG+EDA+IMU	GB	83.33	84.21	63.80 ± 7.60
ECG+EDA+IMU+Tasks	LR	83.33	83.78	65.72 ± 8.49
EDA+IMU+Tasks	NN	83.33	84.21	66.07 ± 5.81
EDA+IMU	LR	83.33	84.21	63.00 ± 4.45
EEG+ECG+EDA	GB	83.33	83.78	59.70 ± 1.16
ECG+EDA+Performance+Tasks	NN	83.33	83.78	70.34 ± 6.92
EEG+ECG+IMU+Performance+Tasks	ET	82.93	82.05	66.47 ± 7.57
EEG+EDA+IMU+Tasks	NN	82.35	84.21	66.00 ± 5.33

Table A.9: 9PM: 10 Best Modalities for Task Interesting Classification.

Best 10 Modalities	Best Classifier	Test F1	Test Accuracy	CV Avg. Validation Accuracy ± SD
EEG+Performance+Tasks	NB	76.92	77.50	52.75 ± 5.31
ECG+IMU+Performance	GP	75.00	74.36	52.90 ± 4.97
ECG+IMU+Tasks	GP	75.00	74.36	52.17 ± 4.74
ECG+IMU+Performance+Tasks	GP	75.00	74.36	52.54 ± 4.90
ECG+IMU	GP	75.00	74.36	52.17 ± 4.74
EEG+Tasks	NB	73.91	70.00	54.61 ± 9.48
EEG+EDA+IMU+Performance+Tasks	QDA	73.68	73.68	55.40 ± 6.80
EEG+EDA+IMU+Performance	QDA	73.68	73.68	56.52 ± 5.79
EEG+EDA+IMU+Tasks	QDA	73.68	73.68	55.77 ± 5.16
EDA+Tasks	KNN	73.33	78.95	67.34 ± 3.23

Table A.10: 9PM: 10 Best Modalities for Physical Tiredness Classification.

Best 10 Modalities	Best Classifier	Test F1	Test Accuracy	CV Avg. Validation Accuracy ± SD
EEG+ECG+IMU	QDA	50.00	74.36	59.10 ± 4.94
EEG+EDA+IMU+Tasks	QDA	50.00	78.95	65.66 ± 4.01
EEG+IMU+Performance	QDA	47.62	72.50	58.00 ± 4.32
EEG+IMU	QDA	47.62	72.50	58.39 ± 1.44
EEG+ECG+EDA+Performance	QDA	47.06	75.68	67.67 ± 2.13
EEG+ECG+EDA+Performance+Tasks	QDA	47.06	75.68	67.67 ± 2.13
ECG+IMU+Performance+Tasks	QDA	46.15	82.05	67.15 ± 5.28
IMU+Tasks	QDA	46.15	82.50	62.75 ± 4.27
ECG+IMU+Tasks	QDA	46.15	82.05	64.96 ± 5.36
EEG+ECG+IMU+Performance+Tasks	QDA	45.45	69.23	59.86 ± 1.88

Table A.11: 9PM: 10 Best Modalities for Distress Classification.

Best 10 Modalities	Best Classifier	Test F1	Test Accuracy	CV Avg. Validation Accuracy ± SD
EDA+IMU+Performance+Tasks	ET	70.00	84.21	63.34 ± 3.90
EEG+ECG+EDA+IMU+Tasks	GP	68.97	75.68	48.85 ± 2.81
EEG+EDA+IMU+Performance+Tasks	GP	66.67	76.32	52.30 ± 5.04
IMU+Tasks	RF	66.67	80.00	65.69 ± 5.81
EEG+EDA+IMU+Performance	GP	66.67	73.68	54.99 ± 6.86
EEG+ECG+Performance	KNN	66.67	79.49	64.39 ± 3.00
EEG+EDA+IMU+Tasks	GP	66.67	73.68	54.60 ± 6.18
ECG+EDA+IMU+Performance	ET	66.67	81.08	62.20 ± 4.71
IMU	ET	66.67	80.00	65.70 ± 2.79
EEG+EDA+IMU	GB	66.67	81.58	63.73 ± 2.22

Table A.12: 9PM: 10 Best Modalities for Attention Classification.

Best 10 Modalities	Best Classifier	Test F1	Test Accuracy	CV Avg. Validation Accuracy ± SD
EDA+IMU+Tasks	AB	70.59	86.84	61.00 ± 7.41
ECG+EDA+IMU+Tasks	AB	64.00	75.68	62.13 ± 6.27
ECG+EDA+IMU+Performance+Tasks	AB	63.16	81.08	62.50 ± 7.74
ECG	RF	62.50	84.62	57.78 ± 4.56
ECG+EDA+IMU	AB	61.54	86.49	57.64 ± 6.61
ECG+IMU+Tasks	NB	60.00	69.23	54.00 ± 7.19
EEG+ECG+IMU+Tasks	KNN	58.82	82.05	63.84 ± 3.80
EEG+ECG+IMU+Performance	KNN	58.82	82.05	63.12 ± 3.69
ECG+IMU+Performance	GB	58.33	74.36	62.04 ± 6.22
ECG+EDA+IMU+Performance	QDA	57.14	75.68	56.56 ± 7.77

Table A.13: 9PM: 10 Best Modalities for Tasks Classification.

Best 10 Modalities	Best Classifier	Test F1	Test Accuracy	CV Avg. Validation Accuracy ± SD
ECG+EDA+Performance	LR	83.37	83.56	71.28 ± 1.57
EEG+EDA+IMU+Performance	RF	80.62	81.08	73.50 ± 1.16
EEG+Performance	GB	80.36	80.36	71.23 ± 4.21
EEG+ECG+Performance	GB	79.91	80.00	71.88 ± 3.78
EEG+EDA+Performance	GB	79.89	79.87	74.27 ± 1.60
EEG+ECG+EDA+IMU+Performance	RF	79.42	80.00	71.21 ± 1.77
EDA+Performance	RF	79.23	79.87	72.89 ± 1.86
EEG+ECG+EDA+Performance	GB	78.47	78.77	73.89 ± 1.73
EDA+IMU+Performance	DT	78.42	79.05	72.98 ± 2.10
ECG+EDA+IMU+Performance	DT	77.97	78.62	72.98 ± 2.10

APPENDIX B

LIST OF PUBLICATIONS

1. Varun Kanal, Maher Abujelala, James Brady, Glenn Wylie, and Fillia Makedon. Adaptive robotic rehabilitation using muscle fatigue as a trigger. In *Proceedings of the 16th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. ACM, 2019

2. Akilesh Rajavenkatanarayanan, Varun Kanal, Konstantinos Tsiakas, Diane Calderon, Michalis Papakostas, Maher Abujelala, Marnim Galib, James C Ford, Glenn Wylie, and Fillia Makedon. A survey of assistive technologies for assessment and rehabilitation of motor impairments in multiple sclerosis. *Multimodal Technologies and Interaction*, 3(1):6, 2019

3. Konstantinos Tsiakas, Maher Abujelala, Akilesh Rajavenkatanarayanan, and Fillia Makedon. User skill assessment using informative interfaces for personalized robot-assisted training. In *International Conference on Learning and Collaboration Technologies*, pages 88–98. Springer, 2018

4. Maher Abujelala, Sanika Gupta, and Fillia Makedon. A collaborative assembly task to assess worker skills in robot manufacturing environments. In *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*, pages 118–119. ACM, 2018

5. Michalis Papakostas, Konstantinos Tsiakas, Maher Abujelala, Morris Bell, and Fillia Makedon. v-cat: A cyberlearning framework for personalized cognitive skill assessment and training. In *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*, pages 570–574. ACM, 2018

6. Konstantinos Tsiakas, Maher Abujelala, and Fillia Makedon. Task engagement as personalization feedback for socially-assistive robots and cognitive training. *Technologies*, 6(2):49, 2018

7. Varun Kanal, Maher Abujelala, Srujana Gattupalli, Vassilis Athitsos, and Fillia Makedon. Apsen: Pre-screening tool for sleep apnea in a home environment. In *International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management*, pages 36–51. Springer, 2017

8. Ashwin Ramesh Babu, Akilesh Rajavenkatanarayanan, Maher Abujelala, and Fillia Makedon. Votre: A vocational training and evaluation system to compare training approaches for the workplace. In *International Conference on Virtual, Augmented and Mixed Reality*, pages 203–214. Springer, 2017

9. Karthikeyan Rajamani, Adhavann Ramalingam, Srinivas Bavisetti, and Maher Abujelala. Cbren: Computer brain entertainment system using neural feedback cognitive enhancement. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments*, pages 236–237. ACM, 2017

10. Alexandros Lioulemes, Michail Theofanidis, Varun Kanal, Konstantinos Tsiakas, Maher Abujelala, Chris Collander, William B Townsend, Angie Boisselle, and Fillia Makedon. Magni dynamics: A vision-based kinematic and dynamic upper-limb model for intelligent robotic rehabilitation. *International Journal of Biomedical and Biological Engineering, World Academy of Science, Engineering and Technology*, 11:158–167, 2017

11. Konstantinos Tsiakas, Cheryl Abellanoza, Maher Abujelala, Michalis Papakostas, Tasnim Makada, and Fillia Makedon. Towards designing a socially assistive robot for adaptive and personalized cognitive training. In

Proceedings of the Robots, volume 4, 2017

12. Konstantinos Tsiakas, Maher Abujelala, Alexandros Lioulemes, and Fillia Makedon. An intelligent interactive learning and adaptation framework for robot-based vocational training. In *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–6. IEEE, 2016

13. Benjamin Chebaa, Alexandros Lioulemes, Maher Abujelala, Dylan Ebert, Scott Phan, Eric Becker, and Fillia Makedon. Multimodal analysis of serious games for cognitive and physiological assessment. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, page 30. ACM, 2016

14. Alexandros Lioulemes, Michalis Papakostas, Shawn N Gieser, Theodora Toutountzi, Maher Abujelala, Sanika Gupta, Christopher Collander, Christoper D Mcmurrough, and Fillia Makedon. A survey of sensing modalities for human activity, behavior, and physiological monitoring. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, page 16. ACM, 2016

15. Maher Abujelala, Cheryl Abellanoza, Aayush Sharma, and Fillia Makedon. Brain-ee: Brain enjoyment evaluation using commercial eeg headband. In *Proceedings of the 9th acm international conference on pervasive technologies related to assistive environments*, page 33. ACM, 2016

16. Maher Abujelala, Alexandros Lioulemes, Paul Sassaman, and Fillia Makedon. Robot-aided rehabilitation using force analysis. In *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, page 97. ACM, 2015

BIBLIOGRAPHY

[1] Maher Abujelala, Cheryl Abellanoza, Aayush Sharma, and Fillia Makedon. Brain-ee: Brain enjoyment evaluation using commercial eeg headband. In *Proceedings of the 9th acm international conference on pervasive technologies related to assistive environments*, page 33. ACM, 2016.

[2] Maher Abujelala, Sanika Gupta, and Fillia Makedon. A collaborative assembly task to assess worker skills in robot manufacturing environments. In *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*, pages 118–119. ACM, 2018.

[3] Maher Abujelala, Alexandros Lioulemes, Paul Sassaman, and Fillia Makedon. Robot-aided rehabilitation using force analysis. In *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, page 97. ACM, 2015.

[4] Mamoun Al-Mardini, Fadi Aloul, Assim Sagahyroon, and Luai Al-Husseini. Classifying obstructive sleep apnea using smartphones. *Journal of biomedical informatics*, 52:251–259, 2014.

[5] JF Alonso, S Romero, MR Ballester, RM Antonijoan, and MA Mañanas. Stress assessment based on eeg univariate features and functional connectivity measures. *Physiological measurement*, 36(7):1351, 2015.

[6] Ashwin Ramesh Babu, Akilesh Rajavenkatanarayanan, Maher Abujelala, and Fillia Makedon. Votre: A vocational training and evaluation system to compare training approaches for the workplace. In *International Conference on Virtual, Augmented and Mixed Reality*, pages 203–214. Springer, 2017.

[7] Alan Baddeley. Working memory. *Current Biology*, 20(4):R136–R140, 2010.

[8] M Safwan Badr. Central sleep apnea. *Primary Care: Clinics in Office Practice*, 32(2):361–374, 2005.

[9] Francisco Barceló, Juan M Muñoz-Céspedes, Miguel A Pozo, and Francisco J Rubia. Attentional set shifting modulates the target p3b response in the wisconsin card sorting test. *Neuropsychologia*, 38(10):1342–1355, 2000.

[10] S Beer, B Aschbacher, D Manoglou, E Gamper, J Kool, and J Kesselring. Robot-assisted gait training in multiple sclerosis: a pilot randomized trial. *Multiple Sclerosis Journal*, 14(2):231–236, 2008.

[11] Biosignalsplux — wearable body sensing platform - Explorer.

[12] Terri Blackwell, Susan Redline, Sonia Ancoli-Israel, Jennifer L Schneider, Susan Surovec, Nathan L Johnson, Jane A Cauley, Katie L Stone, and Study of Osteoporotic Fractures Research Group. Comparison of sleep parameters from actigraphy and polysomnography in older women: the sof study. *Sleep*, 31(2):283–291, 2008.

[13] Laura Bonzano, Andrea Tacchino, Giampaolo Brichetto, Luca Roccatagliata, Adriano Dessypris, Paola Feraco, Maria L Lopes De Carvalho, Mario A Battaglia, Giovanni L Mancardi, and Marco Bove. Upper limb motor rehabilitation impacts white matter microstructure in multiple sclerosis. *Neuroimage*, 90:107–116, 2014.

[14] Alfredo Burgos, Alfredo Goñi, Arantza Illarramendi, and Jesús Bermúdez. Real-time detection of apneas on a pda. *IEEE Transactions on Information Technology in Biomedicine*, 14(4):995–1002, 2009.

[15] Ian G Campbell. Eeg recording and analysis for sleep research. *Current protocols in neuroscience*, 49(1):10–2, 2009.

[16] Raymundo Cassani, Hubert Banville, and Tiago H Falk. Mules: An open source eeg acquisition and streaming server for quick and simple prototyping and recording. In *Proceedings of the 20th International Conference on Intelligent User Interfaces Companion*, pages 9–12. ACM, 2015.

[17] Trudie Chalder, G Berelowitz, Teresa Pawlikowska, Louise Watts, S Wessely, D Wright, and EP Wallace. Development of a fatigue scale. *Journal of psychosomatic research*, 37(2):147–153, 1993.

[18] John Chalmers. *Clinician's Manual on Blood Pressure & Stroke Prevention*. Science Press, 2000.

[19] Guillaume Chanel, Cyril Rebetez, Mireille Bétrancourt, and Thierry Pun. Boredom, engagement and anxiety as indicators for adaptation to difficulty in games. In *Proceedings of the 12th international conference on Entertainment and media in the ubiquitous era*, pages 13–17. ACM, 2008.

[20] Benjamin Chebaa, Alexandros Lioulemes, Maher Abujelala, Dylan Ebert, Scott Phan, Eric Becker, and Fillia Makedon. Multimodal analysis of serious games for cognitive and physiological assessment. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, page 30. ACM, 2016.

[21] Shawn E Christ, Daniel D Holt, Desirée A White, and Leonard Green. Inhibitory control in children with autism spectrum disorder. *Journal of autism and developmental disorders*, 37(6):1155–1165, 2007.

[22] Mario Cifrek, Vladimir Medved, Stanko Tonković, and Saša Ostojić. Surface emg based muscle fatigue evaluation in biomechanics. *Clinical Biomechanics*, 24(4):327–340, 2009.

[23] ApneaLink Air — ResMed.

[24] ARES^{TM} Home Sleep Test – SleepMed — Better Sleep. Better Health.

[25] How Much Does a Sleep Study Cost? (Rates, Fees, & Discounts).

[26] Peter E Comalli Jr, Seymour Wapner, and Heinz Werner. Interference effects of stroop color-word test in childhood, adulthood, and aging. *The Journal of genetic psychology*, 100(1):47–53, 1962.

[27] PLM Coorevits, LA Danneels, Herman Ramon, Remy Van Audekercke, DC Cambier, and GG Vanderstraeten. Statistical modelling of fatigue-related electromyographic median frequency characteristics of back and hip muscles during a standardized isometric back extension test. *Journal of Electromyography and Kinesiology*, 15(5):444–451, 2005.

[28] Trigno^{TM} avanti platform, Apr 2019.

[29] Douglas R Denney and Sharon G Lynch. The impact of multiple sclerosis on patients' performance on the stroop test: processing speed versus interference. *Journal of the International Neuropsychological Society*, 15(3):451–458, 2009.

[30] Sureyya S Dikmen, Patricia J Bauer, Sandra Weintraub, Dan Mungas, Jerry Slotkin, Jennifer L Beaumont, Richard Gershon, Nancy R Temkin, and Robert K Heaton. Measuring episodic memory across the lifespan: Nih toolbox picture sequence memory test. *Journal of the International Neuropsychological Society*, 20(6):611–619, 2014.

[31] Gammon M Earhart, Jim T Cavanaugh, Terry Ellis, Matt P Ford, K Bo Foreman, and Lee Dibble. The 9-hole peg test of upper extremity func-

tion: average values, test-retest reliability, and factors contributing to performance in people with parkinson disease. *Journal of Neurologic Physical Therapy*, 35(4):157–163, 2011.

[32] Danny J Eckert and Atul Malhotra. Pathophysiology of adult obstructive sleep apnea. *Proceedings of the American thoracic society*, 5(2):144–153, 2008.

[33] Michael D. Ellis, Theresa Sukal-Moulton, and Julius P. A. Dewald. Progressive shoulder abduction loading is a crucial element of arm rehabilitation in chronic stroke. *Neurorehabilitation and Neural Repair*, 23(8):862–869, 2009. PMID: 19454622.

[34] Basics: Eyetribe docs.

[35] Peter Feys, Ilse Lamers, Gordon Francis, Ralph Benedict, Glenn Phillips, Nicholas LaRocca, Lynn D Hudson, Richard Rudick, and Multiple Sclerosis Outcome Assessments Consortium. The nine-hole peg test as a manual dexterity performance measure for multiple sclerosis. *Multiple Sclerosis Journal*, 23(5):711–720, 2017.

[36] Leo Galway, Paul McCullagh, Gaye Lightbody, Chris Brennan, and David Trainor. The potential of the brain-computer interface for learning: a technology review. In *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, pages 1554–1559. IEEE, 2015.

[37] Richard C Gershon, David Cella, Nathan A Fox, Richard J Havlik, Hugh C Hendrie, and Molly V Wagster. Assessment of neurological and behavioural function: the nih toolbox. *The Lancet Neurology*, 2010.

[38] Alan S Go, D Mozaffarian, VL Roger, EJ Benjamin, JD Berry, WB Borden, DM Bravata, S Dai, ES Ford, CS Fox, et al. On behalf of the american heart association statistics committee and stroke statistics subcommittee. *Heart disease and stroke statistics—2013 update: a report from the American Heart Association. Circulation*, 127(1):e1–e240, 2013.

[39] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.

[40] Murielle Grangeon, Aymeric Guillot, Pierre-Olivier Sancho, Marion Picot, Patrice Revol, Gilles Rode, and Christian Collet. Rehabilitation of the elbow extension with motor imagery in a patient with quadriplegia after tendon transfer. *Archives of Physical Medicine and Rehabilitation*, 91(7):1143 – 1146, 2010.

[41] David A Grant and Esta Berg. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a weigl-type card-sorting problem. *Journal of experimental psychology*, 38(4):404, 1948.

[42] Jeffrey A Greenberg, John F Burke, Rafi Haque, Michael J Kahana, and Kareem A Zaghloul. Decreases in theta and increases in high frequency activity underlie associative memory encoding. *Neuroimage*, 114:257–263, 2015.

[43] C Guilleminault, R Korobkin, and R Winkle. A review of 50 children with obstructive sleep apnea syndrome. *Lung*, 159(1):275–287, 1981.

[44] Sandra G Hart. Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 50(9), pages 904–908. Sage publications Sage CA: Los Angeles, CA, 2006.

[45] Susan Homack and Cynthia A Riccio. A meta-analysis of the sensitivity and specificity of the stroop color and word test with children. *Archives of clinical Neuropsychology*, 19(6):725–743, 2004.

[46] Nicole Hurless, Aldijana Mekic, Sebastian Peña, Ethan Humphries, Hunter Gentry, and D Nichols. Music genre preference and tempo alter alpha and beta waves in human non-musicians. *Impulse*, 22(4):1–11, 2013.

[47] George W Hynd, Margaret Semrud-Clikeman, Alison R Lorys, Edward S Novey, and Deborah Eliopulos. Brain morphology in developmental dyslexia and attention deficit disorder/hyperactivity. *Archives of neurology*, 47(8):919–926, 1990.

[48] Ekaterina Ivanova, Jörg Krüger, Robert Steingräber, Simone Schmid, Henning Schmidt, and Stefan Hesse. Design and concept of a haptic robotic telerehabilitation system for upper limb movement training after stroke. In *Rehabilitation Robotics (ICORR), 2015 IEEE International Conference on*, pages 666–671. IEEE, 2015.

[49] Craig Jackson. The chalder fatigue scale (cfq 11). *Occupational medicine*, 65(1):86–86, 2014.

[50] Arthur R Jensen. Scoring the stroop test. *Acta psychologica*, 24(5):398–408, 1965.

[51] Susan Jensen and Barbara A Given. Fatigue affecting family caregivers of cancer patients. *Cancer Nursing*, 14(4):181–187, 1991.

[52] Sung-Kwang Ju and Won-Gyu Yoo. Electromyography activity of the deltoid muscle of the weight-bearing side during shoulder flexion in various weight-bearing positions. *Journal of physical therapy science*, 27(10):3285—3286, October 2015.

[53] Sung-kwang Ju and Won-gyu Yoo. Electromyography activity of the deltoid muscle of the weight-bearing side during shoulder flexion in various weight-bearing positions. *Journal of physical therapy science*, 27(10):3285–3286, 2015.

[54] Varun Kanal, Maher Abujelala, James Brady, Glenn Wylie, and Fillia Makedon. Adaptive robotic rehabilitation using muscle fatigue as a trigger. In *Proceedings of the 16th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. ACM, 2019.

[55] Varun Kanal, Maher Abujelala, Srujana Gattupalli, Vassilis Athitsos, and Fillia Makedon. Apsen: Pre-screening tool for sleep apnea in a home environment. In *International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management*, pages 36–51. Springer, 2017.

[56] Thrasyvoulos Karydis, Filipe Aguiar, Simmie L Foster, and Andreas Mershin. Performance characterization of self-calibrating protocols for wearable eeg applications. In *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, page 38. ACM, 2015.

[57] Thrasyvoulos Karydis, Filipe Aguiar, Simmie L Foster, and Andreas Mershin. Performance characterization of self-calibrating protocols for wearable eeg applications. In *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, page 38. ACM, 2015.

[58] Erica Kholinne, Rizki Fajar Zulkarnain, Yu Cheng Sun, SungJoon Lim, Jae-Myeung Chun, and In-Ho Jeon. The different role of each head of the triceps brachii muscle in elbow extension. *Acta orthopaedica et traumatologica turcica*, 2018.

[59] Kinect for windows.

[60] Won Lee, Swamy Nagubadi, Meir H Kryger, and Babak Mokhlesi. Epidemiology of obstructive sleep apnea: a population-based perspective. *Expert review of respiratory medicine*, 2(3):349–364, 2008.

[61] You-Yun Lee and Shulan Hsieh. Classifying different emotional states by means of eeg-based functional connectivity patterns. *PloS one*, 9(4):e95415, 2014.

[62] Gang Li, Boon-Leng Lee, and Wan-Young Chung. Smartwatch-based wearable eeg system for driver drowsiness detection. *IEEE Sensors Journal*, 15(12):7169–7180, 2015.

[63] Zhen Li, Jianjun Xu, and Tingshao Zhu. Prediction of brain states of concentration and relaxation in real time with portable electroencephalographs. *arXiv preprint arXiv:1509.07642*, 2015.

[64] Alexandros Lioulemes, Michalis Papakostas, Shawn N Gieser, Theodora Toutountzi, Maher Abujelala, Sanika Gupta, Christopher Collander,

Christoper D Mcmurrough, and Fillia Makedon. A survey of sensing modalities for human activity, behavior, and physiological monitoring. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, page 16. ACM, 2016.

[65] Alexandros Lioulemes, Paul Sassaman, Shawn N. Gieser, Vangelis Karkaletsis, Fillia Makedon, and Vangelis Metsis. Self-managed patient-game interaction using the barrett WAM arm for motion analysis. In *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments, PETRA 2015, Corfu, Greece, July 1-3, 2015*, pages 34:1–34:8, 2015.

[66] Alexandros Lioulemes, Michail Theofanidis, Varun Kanal, Konstantinos Tsiakas, Maher Abujelala, Chris Collander, William B Townsend, Angie Boisselle, and Fillia Makedon. Magni dynamics: A vision-based kinematic and dynamic upper-limb model for intelligent robotic rehabilitation. *International Journal of Biomedical and Biological Engineering, World Academy of Science, Engineering and Technology*, 11:158–167, 2017.

[67] J Liu, RE Hughes, WP Smutz, G Niebur, and K Nan-An. Roles of deltoid and rotator cuff muscles in shoulder elevation. *Clinical Biomechanics*, 12(1):32–38, 1997.

[68] Rosan Luijcks, Hermie J Hermens, Lonneke Bodar, Catherine J Vossen, Jim van Os, and Richel Lousberg. Experimentally induced stress validated by emg activity. *PloS one*, 9(4):e95215, 2014.

[69] Rosan Luijcks, Catherine J Vossen, Hermie J Hermens, Jim van Os, and Richel Lousberg. The influence of perceived stress on cortical reactivity: a proof-of-principle study. *PloS one*, 10(6):e0129220, 2015.

[70] Peter S Lum, Charles G Burgar, Peggy C Shor, Matra Majmundar, and Machiel Van der Loos. Robot-assisted movement training compared with conventional therapy techniques for the rehabilitation of upper-limb motor function after stroke. *Archives of physical medicine and rehabilitation*, 83(7):952–959, 2002.

[71] Colin M. MacLeod. The Stroop task: The ”gold standard” of attentional measures. *Journal of Experimental Psychology: General*, 121(1):12–14, 1992.

[72] Colin M MacLeod and Penny A MacDonald. Interdimensional interference in the stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in cognitive sciences*, 4(10):383–391, 2000.

[73] Steve Mann, Ryan Janzen, Hang Wu, Max Hao Lu, and Nitin Guleria. Bright ideas: A wearable interactive” inventometer”(brainwave-based idea display). In *2015 IEEE Games Entertainment Media Conference (GEM)*, pages 1–8. IEEE, 2015.

[74] Virgil Mathiowetz, Karen Weber, Nancy Kashman, and Gloria Volland. Adult norms for the nine hole peg test of finger dexterity. *The Occupational Therapy Journal of Research*, 5(1):24–38, 1985.

[75] Sebastiaan Mathôt, Jean-Baptiste Melmi, Lotje van der Linden, and Stefan Van der Stigchel. The mind-writing pupil: A human-computer interface based on decoding of covert attention through pupillometry. Technical report, PeerJ PrePrints, 2015.

[76] Sebastiaan Mathôt, Elle Van Heusden, and Stefan Van der Stigchel. Attending and inhibiting stimuli that match the contents of visual working

memory: Evidence from eye movements and pupillometry. Technical report, PeerJ PrePrints, 2015.

[77] Gerald Matthews, James Szalma, April Rose Panganiban, Catherine Neubauer, and Joel S Warm. Profiling task stress with the dundee stress state questionnaire. *Psychology of stress: New research*, 1:49–90, 2013.

[78] Writing Group Members, Véronique L Roger, Alan S Go, Donald M Lloyd-Jones, Emelia J Benjamin, Jarett D Berry, William B Borden, Dawn M Bravata, Shifan Dai, Earl S Ford, et al. Executive summary: heart disease and stroke statistics—2012 update: a report from the american heart association. *Circulation*, 125(1):188–197, 2012.

[79] Alma S Merians, David Jack, Rares Boian, Marilyn Tremaine, Grigore C Burdea, Sergei V Adamovich, Michael Recce, and Howard Poizner. Virtual reality–augmented rehabilitation for patients following stroke. *Physical therapy*, 82(9):898–915, 2002.

[80] MetaMotionR - MbientLab.

[81] Microsoft band 2 features and functions.

[82] Maddalena Mugnosso, Francesca Marini, Matteo Gillardo, Pietro Morasso, and Jacopo Zenzeri. A novel method for muscle fatigue assessment during robot-based tracking tasks. In *Rehabilitation Robotics (ICORR), 2017 International Conference on*, pages 84–89. IEEE, 2017.

[83] Eeg and biofeedback research from muse.

[84] 9-Hole Peg Test (9-HPT) : National Multiple Sclerosis Society.

[85] Bibhukalyan Prasad Nayak, Sibsambhu Kar, Aurobinda Routray, and Akhaya Kumar Padhi. A biomedical approach to retrieve information on driver's fatigue by integrating eeg, ecg and blood biomarkers during simulated driving session. In *Intelligent Human Computer Interaction (IHCI), 2012 4th International Conference on*, pages 1–6. IEEE, 2012.

[86] Intro to NIH Toolbox.

[87] Pulse oximetry.

[88] Johanna Renny Octavia, Peter Feys, and Karin Coninx. Development of activity-related muscle fatigue during robot-mediated upper limb rehabilitation training in persons with multiple sclerosis: a pilot trial. *Multiple sclerosis international*, 2015, 2015.

[89] Ultracortex Mark IV — OpenBCI Documentation.

[90] Michalis Papakostas, Konstantinos Tsiakas, Maher Abujelala, Morris Bell, and Fillia Makedon. v-cat: A cyberlearning framework for personalized cognitive skill assessment and training. In *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*, pages 570–574. ACM, 2018.

[91] Thomas Penzel, George B Moody, Roger G Mark, Ary L Goldberger, and J Hermann Peter. The apnea-ecg database. In *Computers in Cardiology 2000. Vol. 27 (Cat. 00CH37163)*, pages 255–258. IEEE, 2000.

[92] Luka Peternel, Nikos Tsagarakis, Darwin Caldwell, and Arash Ajoudani. Robot adaptation to human physical fatigue in human–robot co-manipulation. *Autonomous Robots*, pages 1–11, 2018.

[93] Scott Phan, Alexandros Lioulemes, Cyril Lutterodt, Fillia Makedon, and Vangelis Metsis. Guided physical therapy through the use of the barrett WAM robotic arm. In *2014 IEEE International Symposium on Haptic, Audio and Visual Environments and Games, HAVE 2014, Richardson, TX, USA, October 10-11, 2014*, pages 24–28, 2014.

[94] Rosalind W Picard. Affective computing for hci. In *HCI (1)*, pages 829–833. Citeseer, 1999.

[95] A Plotnikov, N Stakheika, Alessandro De Gloria, Carlotta Schatten, Francesco Bellotti, Riccardo Berta, C Fiorini, and Flavio Ansovini. Exploiting real-time eeg analysis for assessing flow in games. In *2012 IEEE 12th International Conference on Advanced Learning Technologies*, pages 688–689. IEEE, 2012.

[96] Janet L Poole, Patricia A Burtner, Theresa A Torres, Cheryl Kirk McMullen, Amy Markham, Michelle Lee Marcum, Jennifer Bradley Anderson, and Clifford Qualls. Measuring dexterity in children using the nine-hole peg test. *Journal of Hand Therapy*, 18(3):348–351, 2005.

[97] Gratianne Rabiller, Ji-Wei He, Yasuo Nishijima, Aaron Wong, and Jialing Liu. Perturbation of brain oscillations after ischemic stroke: a potential biomarker for post-stroke function and therapy. *International journal of molecular sciences*, 16(10):25605–25640, 2015.

[98] Karthikeyan Rajamani, Adhavann Ramalingam, Srinivas Bavisetti, and Maher Abujelala. Cbren: Computer brain entertainment system using neural feedback cognitive enhancement. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments*, pages 236–237. ACM, 2017.

[99] Akilesh Rajavenkatanarayanan, Varun Kanal, Konstantinos Tsiakas, Diane Calderon, Michalis Papakostas, Maher Abujelala, Marnim Galib, James C Ford, Glenn Wylie, and Fillia Makedon. A survey of assistive technologies for assessment and rehabilitation of motor impairments in multiple sclerosis. *Multimodal Technologies and Interaction*, 3(1):6, 2019.

[100] Bruce Ratner.

[101] Paula Rego, Pedro Miguel Moreira, and Luis Paulo Reis. Serious games for rehabilitation: A survey and a classification towards a taxonomy. In *Information Systems and Technologies (CISTI), 2010 5th Iberian Conference on*, pages 1–6. IEEE, 2010.

[102] Mouna Safi-Harb, Sébastien Proulx, Nicolas von Ellenrieder, and Jean Gotman. Advantages and disadvantages of a fast fmri sequence in the context of eeg–fmri investigation of epilepsy patients: a realistic simulation study. *NeuroImage*, 119:20–32, 2015.

[103] Lizawati Salahuddin, Jaegeol Cho, Myeong Gi Jeong, and Desok Kim. Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. In *2007 29th annual international conference of the ieee engineering in medicine and biology society*, pages 4656–4659. IEEE, 2007.

[104] I 1 Sánchez-Cubillo, JA Perianez, D Adrover-Roig, JM Rodriguez-Sanchez, M Rios-Lago, JEEA Tirapu, and F Barcelo. Construct validity of the trail making test: role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *Journal of the International Neuropsychological Society*, 15(3):438–450, 2009.

[105] Scikit-learn: Randomizedsearchcv¶.

[106] Cornelia Setz, Bert Arnrich, Johannes Schumm, Roberto La Marca, Gerhard Tröster, and Ulrike Ehlert. Discriminating stress from cognitive load using a wearable eda device. *IEEE Transactions on information technology in biomedicine*, 14(2):410–417, 2009.

[107] Surgery - surgical procedures.

[108] J Ridley Stroop. Studies of interference in serial verbal reactions. *Journal of experimental psychology*, 18(6):643, 1935.

[109] J Ridley Stroop. Studies of interference in serial verbal reactions. *Journal of experimental psychology*, 18(6):643, 1935.

[110] Abdulhamit Subasi and M Kemal Kiymik. Muscle fatigue detection in emg using time–frequency methods, ica and neural networks. *Journal of medical systems*, 34(4):777–785, 2010.

[111] Leslie M Swanson, J Todd Arnedt, Mark R Rosekind, Gregory Belenky, Thomas J Balkin, and Christopher Drake. Sleep disorders and work performance: findings from the 2008 national sleep foundation sleep in america poll. *Journal of sleep research*, 20(3):487–494, 2011.

[112] Maciej Szaniawski, Stella Boess, Gerald Kraan, and Richard Goossens. Enhancing engagement with motor rehabilitation therapy through gamified rehabilitation interactions. In *Proceedings of the 3rd European Conference on Design 4 Health 2015*, 2015.

[113] Adriana Tapus, Cristian Tapus, and Maja J Mataric. The use of socially assistive robots in the design of intelligent cognitive therapies for people with dementia. In *2009 IEEE international conference on rehabilitation robotics*, pages 924–929. IEEE, 2009.

[114] N Tellez, J Rio, M Tintore, C Nos, I Galan, and X Montalban. Does the modified fatigue impact scale offer a more comprehensive assessment of fatigue in ms? *Multiple Sclerosis Journal*, 11(2):198–202, 2005.

[115] Lucas J Tiesinga, Theo WN Dassen, and Ruud JG Halfens. Fatigue: a summary of the definitions, dimensions, and indicators. *International Journal of Nursing Terminologies and Classifications*, 7(2):51–62, 1996.

[116] Willis H Tsai, W Ward Flemons, William A Whitelaw, and John E Remmers. A comparison of apnea–hypopnea indices derived from different definitions of hypopnea. *American journal of respiratory and critical care medicine*, 159(1):43–48, 1999.

[117] Konstantinos Tsiakas, Cheryl Abellanoza, Maher Abujelala, Michalis Papakostas, Tasnim Makada, and Fillia Makedon. Towards designing a socially assistive robot for adaptive and personalized cognitive training. In *Proceedings of the Robots*, volume 4, 2017.

[118] Konstantinos Tsiakas, Maher Abujelala, Alexandros Lioulemes, and Fillia Makedon. An intelligent interactive learning and adaptation framework for robot-based vocational training. In *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–6. IEEE, 2016.

[119] Konstantinos Tsiakas, Maher Abujelala, and Fillia Makedon. Task engagement as personalization feedback for socially-assistive robots and cognitive training. *Technologies*, 6(2):49, 2018.

[120] Konstantinos Tsiakas, Maher Abujelala, Akilesh Rajavenkatanarayanan, and Fillia Makedon. User skill assessment using informative interfaces for

personalized robot-assisted training. In *International Conference on Learning and Collaboration Technologies*, pages 88–98. Springer, 2018.

[121] Ying-Chih Wang, Richard W Bohannon, Jay Kapellusch, Arun Garg, and Richard C Gershon. Dexterity as measured with the 9-hole peg test (9-hpt) across the age span. *Journal of Hand Therapy*, 28(1):53–60, 2015.

[122] David Watson and Lee Anna Clark. The panas-x: Manual for the positive and negative affect schedule-expanded form, 1999.

[123] Nathaniel F Watson. Health care savings: the economic value of diagnostic and therapeutic care for obstructive sleep apnea. *Journal of Clinical Sleep Medicine*, 12(08):1075–1077, 2016.

[124] Terry Young, Laurel Finn, Paul E Peppard, Mariana Szklo-Coxe, Diane Austin, F Javier Nieto, Robin Stubbs, and K Mae Hla. Sleep disordered breathing and mortality: eighteen-year follow-up of the wisconsin sleep cohort. *Sleep*, 31(8):1071–1078, 2008.