AUTONOMIC NERVOUS SYSTEM APPROACH TO MEASURE PHYSIOLOGICAL AROUSAL AND ASSESS USER EXPERIENCE IN SIMULATION-BASED EMERGENCY PROCEDURE TRAINING ENVIRONMENT

by © Sinh Bui

A Thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of

Master of Engineering

Faculty of Engineering and Applied Science

Memorial University of Newfoundland

October 2018

St. John's Newfoundland and Labrador

Abstract

Given of the impossibility of exposing trainees to hazardous scenarios for ethical, financial and logistical reasons, virtual-environment (VE) based simulation training has been adopted in various safety-critical industries. Through simulation, participants can be exposed to a variety of training scenarios to assess their performance under different conditions. Along with performance measures, physiological signals may provide useful information about trainees' experience. The objective of this research is to investigate the ability of physiological measurement to provide information on trainees' experiences by assessing their physiological arousals in a simulation-based training environment.

In this study, 38 participants used a VE-based program called AVERT (All-hands Virtual Emergency Response Trainer). This program was developed for training emergency response procedures for the offshore petroleum industry. Signals of the autonomic nervous system (ANS), specifically electrocardiography (ECG), electrodermal activities (EDA), and respiration (RSP), were used to assess physiological arousal levels for 8 different conditions of an emergency evacuation task.

On average, neutral and training conditions could be distinguished with an 82.4% average accuracy by a subject-specific machine learning classifier. Most importantly, arousal levels in different training scenarios provide useful information that performance measures alone do not reveal.

Keywords: Human Factors · Physiological Signal · Machine Learning · Virtual Environment · Emergency Response Training · Training Design

Acknowledgements

I would like to extend my sincere thanks to my supervisor – Dr. Sarah Power and co-supervisor – Dr. Brian Veitch, for helping me through the Master program. They both provided invaluable advice that helped me during my studies. This thesis would net have been possible without their support.

I would also like to thank NSERC and Husky Energy Industrial Research Chair in Safety at Sea for their financial support.

Furthermore, a thanks to everyone in my research group. We have collaborated perfectly over the last two years and it has been my honor to be a part of the team. I learned a lot from those talented and humorous people, and also got help from them.

Thanks Dr. Leonard Lye – Associate Dean of the Faculty of Engineering and Applied Science, and Mr. Andrew Kim – Dean of School of Graduate Studies, for their great support before and during the time I did my program at Memorial University of Newfoundland.

Last but not least, I would like to thank my family and friends, who have always been by my side and have shared both the happiness and bitterness in life. They have played a large role in the completion of my degree, for which I am greatly appreciative.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AVERT	All-hands Virtual Emergency Response Trainer
VE	Virtual Environment
PSF	Performance Shaping Factor
ANS	Autonomic nervous system
ECG	Electrocardiography
EDA	Electrodermal Activities
RSP	Respiration Rate
EEG	Electroencephalography
EMG	Electromyography
GSR	Galvanic Skin Response
LDA	Linear Discriminant Analysis
QDA	Quadratic Discriminant Analysis
SVM	Support Vector Machine
HRV	Heart Rate Variability
GPA	General Platform Alarm
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
PCA	Principle Component Analysis
LO	Learning Objective

Chapter 1: Introduction

1.1 Relevance of work

On-site workers in a variety of disciplines are exposed to hazardous environments, where quick and incisive decisions must be made in the event that an emergency situation arises. Therefore, it is critical for such personnel to receive comprehensive training in the proper performance of emergency response procedures. However, exposing workers to realistic emergency conditions for the purposes of training is impossible for ethical, financial, and logistical reasons. Therefore, virtual environments (VE) are often used to simulate safety-critical conditions, allowing participants to experience those situations without actually being exposed to hazards. It is important that training scenarios be designed to have different difficulty levels in order to help trainees acquire skills effectively, as well as confirm their ability to respond to a variety of situations. The method of scoring trainees' performance through a rubric has been employed to assess the difficulty of a particular training scenario (Smith & Veitch, 2015). However, this approach is not comprehensive because it only assesses the number and type of errors the individual performs during the scenario and ignores their cognitive and affective experience of the scenario. For example, a participant could achieve similar performance scores in two different scenarios, but could have experienced significantly different levels of stress and/or mental workload while performing them. This would indicate a difference in difficulty level between the scenarios that performance levels alone would not reveal.

Physiological signal changes - particularly those of the autonomic nervous system (ANS), including electrocardiography, electrodermal activities, and respiration - have been shown to be good indicators of mental workload and stress in a variety of domains (Sharma & Gedeon, 2012). In addition, physiological signal arousal could also be used to assess users' engagement during an activity, especially in VR-based platforms. This capability is possible as a result of the physiological signal changes that occur when an individual is either concentrating on what he/she is doing, or being stressed by a simulated safety-critical situation (Vrijkotte, et al., 2000).

Electrocardiography (ECG)

ECG is the process of recording electrical activity of the heart over a period of time using electrodes placed on the skin. These electrodes detect electrical changes on the skin that arise from the heart muscle's electrophysiologic pattern of depolarizing during each heartbeat. It is a very commonly performed cardiology test. As shown in Fig. 1, there are a variety of features that can be observed in an ECG signal. In this work, we focused on QRS complexes, which provide information about the heart beat. Heart rate can be determined from the time between successive QRS complexes. Heart rate variability (HRV) can also be extracted (Malik, et al., 1996), which is a combination of a number of features in both the time and frequency domains. Generally, with increased emotional arousal, the heart rate tends to increase, and heart rate variability decreases (Vrijkotte, et al., 2000).

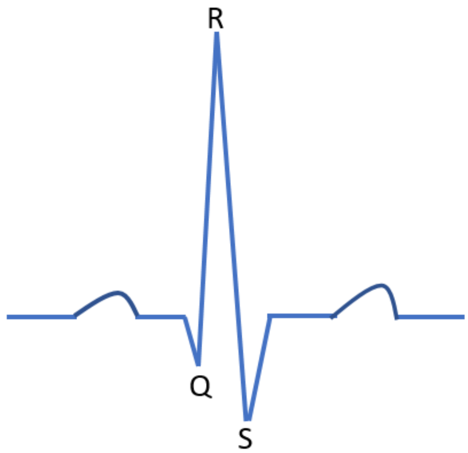


Figure 1. A single QRS complex from a normal ECG, representing one heart beat.

Electrodermal activity (EDA)

Electrodermal activity is a measure of skin conductance. It is affected by sweat gland activity, which is controlled by the sympathetic nervous system – a branch of the autonomous nervous system. If the sympathetic nervous system is aroused, there is an increase in sweat gland activity, leading to higher skin conductance (Conesa, 1995; Carlson, 2013; Figner & Murphy, 2010; Nagai, et al., 2004; Loggia, et al., 2011). Therefore, EDA is an indication of emotional arousal.

Respiration (RSP)

Respiration (RSP) signal tracks the movement of the ribcage over time due to breathing. From this signal, respiration rate (RR) can be extracted, which indicates how fast or slow one is breathing. Literature has shown that our mental states have an effect on the way we respire (Tovian, et al., 2018). Specifically, stress can make us breath harder or might cause rapid breathing.

VE-based simulation training

In terms of VE-based simulation training, these measures have been applied to investigate how immersive a VE is, or the effectiveness of a training simulation in helping trainees better cope with the real situations in terms of mental state. In this study, we investigate the use of autonomic nervous system signals in providing information on the user experience in various emergency response training scenarios completed by naive trainees (i.e., trainees who had neither known or experienced the experiment, nor been offshore on a vessel) in a simulation-based training environment. The study was conducted using a VE-based program called AVERT (All-hands Virtual Emergency Response Trainer) (House, et al., 2014), which was developed for training emergency evacuation procedures for the offshore petroleum industry. AVERT is a software-based training program, where participants use a standard video game controller to direct an avatar through emergency evacuation procedures on a realistic offshore oil platform under various conditions (Figure 2).

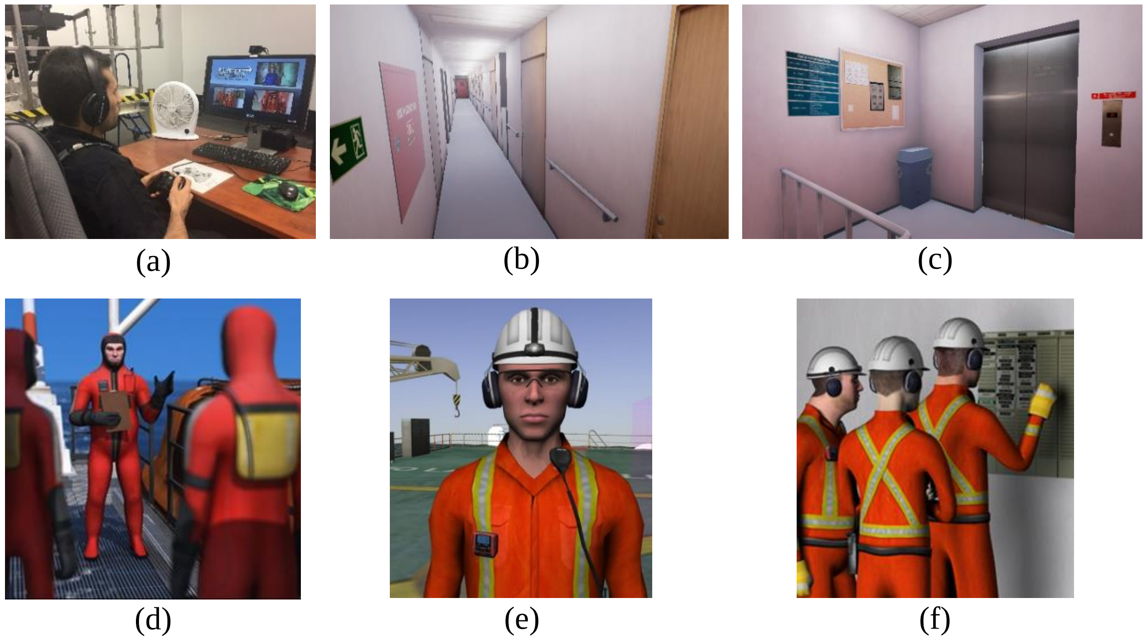


Figure 2. AVERT training application: (a) participant is using the software (b)(c) simulated spaces on a vessel (d)(e)(f) simulated subjects on a vessel

1.2 Objectives

The purposes of this work are to:

1. Use physiological signals to assess stress levels induced by different scenarios in AVERT
2. Use physiological signals to assess which performance shaping factors (PSFs) significantly affect stress levels during different scenarios in a virtual training environment
3. Investigate whether there is any correlation between stress level and performance in the training scenarios
4. Investigate whether there is any correlation between objective and subjective measures of participants' stress during training session

1.3 Thesis structure

Following the introduction, the literature review is presented to outline existing work in the field. This is followed by the Methods section, where detailed information regarding how the study was designed and implemented is described. Thereafter, the Results and Discussion are provided, before the key Conclusions are presented in the last section.

Chapter 2: Literature review

This literature review will discuss the main aspects of applying physiological signals for stress detection in virtual training environments. An overview of research in the field will be provided, including two main points:

1. The advantages, developments and challenges of virtual training environments for simulated hazardous workplace training.
2. The assessment of human stress by physiological measures and its applications in VE-based training platforms.

In this chapter, insight into the virtual training field is first presented, including benefits, developments, and the challenges it is currently facing. This is followed by a brief summary of studies about ANS signals. Research on detecting stress from physiological signals is then outlined alongside a discussion of current applications.

2.1 Virtual reality-based training: advantages, developments, and challenges

Virtual training has proved its capability of providing users with a simulated hazardous environment to practice skills needed to stay safe while working in the real workplace. It is necessary to train personnel to respond to emergency conditions because in these situations it is critical to take appropriate actions as quickly as possible under a time constraint and under high stress conditions (Jones, et al., 1981; Rosenbaum, et al., 1981; Merién, et al., 2010; Chang, et al., 2010; Mantovani, et al., 2003). In order to do this, people need to master the necessary skills and be able to use these skills comfortably even under high stress conditions. Because lecture-based training does not offer realism, and real-life practice is costly and might impose dangers, virtual reality-based training comes into play as a potential alternative that balances realism and cost-effectiveness (Hsu, et al., 2013). Specifically, VR-based programs provide trainees with more interactive scenarios compared to lecture-based method, and these programs are less expensive than real-life practice. In addition, VR-based training offers adaptiveness to various industries, consistency in different trials, incorporation of varied stimuli, instant feedback, individual and group-orientation, the ability to create complicated scenarios, and an effective evaluation process (Seymour, et al., 2002; Gurusamy, et al., 2009; Gallagher, et al., 1999; Ahlberg, et al., 2007; Mantovani, et al., 2003; Aoki, et al., 2007; Andreatta, et al., 2010).

Because of its assets, virtual training has been adopted in a variety of industries, especially those involving hazardous work places like nuclear power plants, offshore oil rigs, spaceships, and airplanes, where one mistake from a worker could lead to severe accidents. For example, Erren-Wolters, et al., (2007) conducted a review on VR-based training applications. Five studies were reviewed in this paper were related to training driving skills, one study was related to physical exercise training and one to leisure activity. The review suggested that VR could be a useful method to improve the control of a mobility device or to keep up the physical condition, or a means of leisure activity. Van Wyk and De Villiers (2009) provided information about VR-based program applied in safety training in mines in Africa, including contextual requirements and difficulties. The study showed that VR-based training helped improving the safety culture and awareness of the mining employees and recommended VR-based training program as a promising method to be applied to the mining industry in Africa. After reviewing a number of papers related to VR applications in simulating medical scenarios, Rizzo, et al., (2009) concluded that VR could provide identical simulation environments where performance can be measured and rehabilitated, which would be helpful in treatment or rehabilitation purposes. In another study, Chan, et al., (2011) proposed a virtual reality dance training system using motion capture technology, where a student can follow a simulated teacher's motions. The results proved that the system can successfully instruct trainees to enhance their skills.

Although applied in various fields, VR-based training platforms are facing challenges that prevent it from becoming more effective (Shaw, et al., 2015; Waycott, et al., 2018; Roth, et al., 2015). Firstly, not everyone is familiar with the virtual reality technology, thus the ability to finish a virtual training session varies among individuals (Hsu, et al., 2013). Secondly, although implementing VR-based training is relatively cost-effective, developing a highly immersive environment still consumes a significant amount of resources. Last, but not least, because of the limitation of technology, there is experience available in real practice that could not be simulated accurately in a VE. Nevertheless, it is believed by experts that technology's continuous development will overcome these challenges (Hsu, et al., 2013). One of the approaches that researchers have investigated and proved the applicability to help overcome these challenges is measuring user's stress/mental work load during virtual training session. The following section will provide more insight into this approach.

2.2 Assessing stress by physiological measures and application in VR-based training

Stress is a state of mental or emotional strain or tension resulting from situations which are difficult and require a significant amount of mental resources. When a person is doing a task, a reasonable amount of stress might result in better performance, because he or she is concentrating. However, if the stress level is too high, the result might be negatively affected (Yerkes & Dodson, 1908). Physiological signal changes - particularly those of the autonomic nervous system (ANS), including electrocardiography, galvanic skin response, and respiration - have been shown to be good indicators of mental workload and stress in a variety of domains (Sharma & Gedeon, 2012). For example, Plarre, et al., (2011) proposed an approach to detect people's daily stress and send an early warning to them, which applies two different models: *physiological classifier* and *perceived stress model.* The first model provides information about the variation in physiological states, while the other uses this information to calculate the probability of stress. The authors applied wearable devices to participants, which provides electrocardiography (ECG) and respiratory inductive plethysmography (RIP). The result reveals that subjects were stressed in 35.14% of time in terms of physiological aspect, while that number is 26.61% and 28.08% in terms of perceived stress model and self-report, respectively. Ultimately, they concluded that there were three main points they had drawn from the study, including correcting the differences between individuals, the fact that respiration features provided high discrimination, and a new model mapping physiological stress to perceived stress. Xu, et al., (2015) suggested a cluster-based method to detect stress level, where they focus on solving the problem of variability in stress response among people. In this study, electroencephalography (EEG), ECG, electromyography (EMG), and galvanic skin response (GSR, also acknowledged as EDA) were employed. They concluded that when using the cluster-based method, the stress detection accuracy increases significantly compared to previous methods without clustering. Ollander, et al., (2016) compared how the stress detection performance of the Empatica E4 wrist band compared to that of stationary sensors, using ECG and EDA signals. Results showed that although there was a noticeable loss in inter-beat intervals, the wrist band retained high accuracy in time-domain features like mean and standard deviation of heart rate, which provides more information regarding levels of stress. In another study, Smets, et al., (2016) aimed to find the most efficient algorithm for detecting stress based on physiological signals. The data that were used were ECG, GSR, skin temperature and respiration rate. Six different machine learning algorithms were applied to the data to detect the stress level. It was found that personalized Bayesian networks and generalized support vector machines derived the highest average stress detection results with 84.6% and 82.7% respectively. Most recently, Huysmans, et al., (2018) proposed a method of applying unsupervised learning, specifically Self-Organizing Maps (SOM) for stress detection. Skin conductance (SC, also acknowledged as EDA) and ECG were used as input data. It was concluded that the SOM-based technique was capable of detecting stress, with a comparable accuracy to previous methods.

Stress detection has been applied in a number of studies to learn more about the VR-based training environment, as well as how trainees performed in terms of mental workload/stress. For example, Patton and Gamble (2016) developed and tested an immersive VE to train soldiers. They measured participants' physiological arousal through heart rate variabilities (HRV) and showed that it could be used as an indicator of the level of stress or immersion the simulation caused. In another study, Lackey, et al., (2016) conducted research to evaluate virtual reality's effectiveness, where they investigated the stress and workload that participants experienced in a real training scenario after being trained virtually. It was found that participants who reported a positive experience in virtual training performed better in real training, which was inferred from lower stress and workload. Meanwhile, Egan, et al., (2016) suggested a measure of heart rate (HR) and EDA as an objective method to assess quality of experience (QoE) for immersive VR environments. Results from this research show that while there was not a significant change in participant's HR between VR and non-VR environments, there was a significant effect of environment on both EDA and subjective ratings, as well as a significant correlation between EDA and subjective ratings results. It was concluded that EDA might be more effective than HR in indicating how one experienced VR environments, and how the VR environment brought to users a different experience compared to the non-VR environment. Bian, et al., (2015) built a VR-based program to train young people with autism spectrum disorder (ASD) to drive. In order to investigate the effects of various feelings (engagement, enjoyment, frustration, and boredom) on performance, physiological signals were collected during training sessions before being analyzed by six different classifiers. It was suggested by the results that the method developed in this study could reliably recognize physiological arousals in teenagers with ASD and provide the basis for physiological-related affect-sensitive driving skill training system.

These studies along with a number of others (Panju, et al., 2015; Garcia-Ceja, et al., 2016; Cho, et al., 2017; Rizzo, et al., 2012; Basdogan, et al., 2001) have shown the possibility of applying physiological measure to access human’s mental workload/stress, particularly in VR-based environments.

Knowledge gap: Although there are studies about performance in virtual emergency training environment, the potential of applying physiological changes detection to assess trainees’ experiences during various virtual emergency training sessions with different difficulty levels remains unknown.

Chapter 3: Methodology

3.1 Experiment design

3.1.1 Participants

Data from 38 participants (28 males and 10 females) were collected during the experiment. Participants' ages ranged from 18 to 65 years old. Participants were excluded if they had any prior experience with the experimental procedure, or with the real-life offshore petroleum platform on which it is based. This exclusion was made to ensure all participants started the learning process from the same position, which provided the objectiveness to the data. The subjects were asked not to consume alcohol within 24 hours of the experimental session. Also they were asked to refrain from exercising, consuming food or caffeinated beverages, and smoking within 2 hours of the experimental session. This was asked of participants to ensure the physiological signals of interest would not be affected[1]. Approval of the experimental protocol was obtained from the appropriate research ethics board at Memorial University of Newfoundland (ICEHR #20171099) prior to study commencement. All participants provided written, informed consent prior to participation.

3.1.2 Experimental protocol

The experiment consisted of one session, divided into two phases. In Phase 1, the participants were first trained to be competent in basic emergency evacuation skills using AVERT. Specifically, they went through a series of modules consisting of instructional

[1] The full recruiting information is provided in Appendix A

material and practice trials to become familiar with tasks such as recognizing various alarms, travelling to an appropriate muster station from their cabin, and selecting appropriate personal protective equipment. Training involved the completion of four scenarios addressing different learning objectives (House, et al., 2014):

(1) Establish spatial awareness of environment

(2) Alarms recognition: understand role of alarms and urgency of situation

(3) Routes and mapping: determine primary and alternative routes to muster stations

(4) Perform muster station protocol and individual responsibility

(5) Safe practices

(6) First actions – taking appropriate equipment from cabin

Participants were required to re-attempt each training scenario until they could complete it error-free before moving on to the next module and training scenario (i.e., a mastery learning approach (Block & Burns, 1976) was taken).

After the participants were trained in the basic evacuation procedures, and demonstrated a minimum level of competence, they were given a short break and then began the second phase of the experiment. In this phase, participants were asked to perform the evacuation procedures they had learned in Phase 1, but this time under various new conditions. A 2^3 factorial design (Montgomery, 2013) was employed: 38 participants completed eight (8) scenarios based on three (3) performance shaping factors (PSFs), each varied at two levels (low and high). A description of the PSFs is provided in Table 1. The aim of this design was to create scenarios with different levels of difficulty.

Before completing each of the eight training scenarios, participants completed a five-minute rest interval in order to give them a break, and to get a baseline measure of their physiological signals. The average duration of one training scenario was 237 s +/- 138 s. Participants had a limit of ten minutes to complete a scenario. There was considerable variation in the duration of scenarios across conditions and participants, the average duration of one training scenario being 237 s +/- 138 s. The limit for a participant to finish a scenario was ten minutes; the scenario ended after this time if the participant was not able to finish. The order in which the scenarios were performed was randomized for each participant. Figure 3 shows the trial sequence for Phase 2 of the experiment. Note that the i^{th} scenario can be any scenario from 1 to 8, the sequence of scenarios to be completed for each participant was randomly generated.

Table 1. Factors and levels in the 2^3 factorial design

PSF	Low level	High level
1) Quality of information received over public announcement (PA) system during the scenario	The PA announcement is clear, concise, and includes all relevant information	The PA announcement is not clear and does not provide sufficient information
2) Proximity to hazard	There is no hazard (e.g., fire, explosion, smoke)	There is close proximity to hazard (e.g., fire, explosion, smoke)
3) Familiarity of environment	Scenario starts in familiar location (i.e., from Phase 1), participants take known route, and there is potential for known re-route	Scenario starts in unfamiliar location, there is potential for re-routing based on acquired information

Table 2. Scenario name and the corresponding PSF levels

Scenario name	PSF 1	PSF 2	PSF 3
1	Low	Low	Low
2	High	Low	Low
3	Low	High	Low
4	Low	Low	High
5	High	High	Low
6	High	Low	High
7	Low	High	High
8	High	High	High

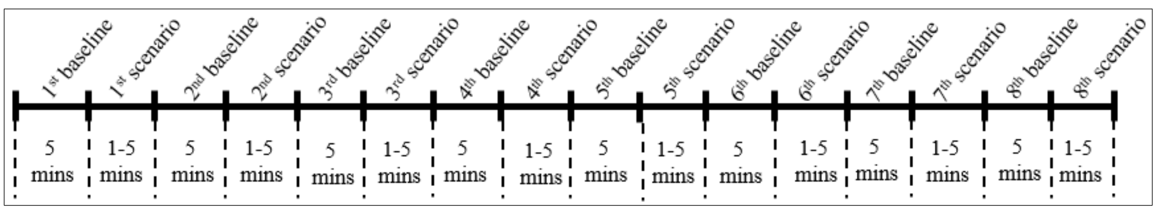


Figure 3. Experiment baseline-scenario sequence

3.2 Performance scoring

To evaluate each participant’s performance, the following information was collected for each scenario:

(1) Alarm recognition: did the participant recognize the meanings of different alarm types and react accordingly?

(2) Identification of mustering announcement: did the participant muster at the correct location and perform the correct task after reaching the muster station (e.g., put on immersion suit)?

(3) Route selection: which route did the participant take in a given situation, and did they re-route appropriately when a hazard (e.g., fire, smoke) was encountered?

(4) Observation of general safety rules: did the participant close all safety doors, and walk, not run, on the platform?

Based on this performance data, a performance score was calculated for each participant in each scenario (see Table 3 for an example of rubric for Scenario 1[2]) (House, et al., 2014).

[2] The other rubrics for the remaining scenarios could be found in Appendix B

Table 3. Performance scoring rubric for Scenario 1

Learning Objectives	Specific Tasks	Performance Measure	Weighting
LO1. Establish Spatial Awareness of Environment	Identify Primary Muster Station	Correct location	See LO2
LO2. Alarms Recognition: Understand role of alarms and urgency of situation	Identify General Platform Alarm (GPA)	Correct location (GPA = Mess Hall, Proper Activity = Lifeboat)	25
LO3. Routes and Mapping: Determine Primary and Alternative Routes to Muster Stations	Accommodation Cabin to Primary Muster Station	Route selected (prim, second, or others) & off route 15 points primary; 7.5 secondary; 0 lost or off route	15
	Primary Muster Station back to Cabin	Route selected (prim, second, or others) & off route	15
LO4. Perform Muster Station Protocol and Individual Responsibilities	Perform T-Card Procedure at Muster Station	Correct location + Move t-card correctly	12.5
		Un-muster	12.5
LO5. Safe Practices	Do not run on the platform	Speed of trainee (% running)	10
	Recognize and Use Fire Doors & Water Tight Doors	Number of fire/water tight doors left open (closed)	15
LO6. First Actions - Taking proper equipment from Cabin	Know to locate and bring the following: Grab Bag and Immersion Suit	Takes Grab Bag and Immersion Suit	10
		Total	115

3.3 Physiological measure

As an indicator of the level of stress or workload experienced by each participant in each of the eight scenarios, the classification accuracy of each scenario versus the baseline interval that preceded it was used. The baseline was assumed to represent a low arousal state (e.g., low stress). Classification accuracy is a measure of the separateness of data, thus the higher the classification accuracy between a scenario and the baseline condition, the more the physiological signals changed during the scenario, indicating a higher arousal state (e.g., high stress/workload). In order to derive this classification accuracy, participant's physiological signals were first recorded during the experiment, then they were preprocessed and useful features were extracted. Thereafter, the feature set dimensionality was reduced by selecting the most discriminatory ones from the full feature set before supervised machine learning algorithms were applied to classify the data between baseline and scenario. Unsupervised machine learning algorithms were also employed to investigate possible patterns in the data (see Figure 4).

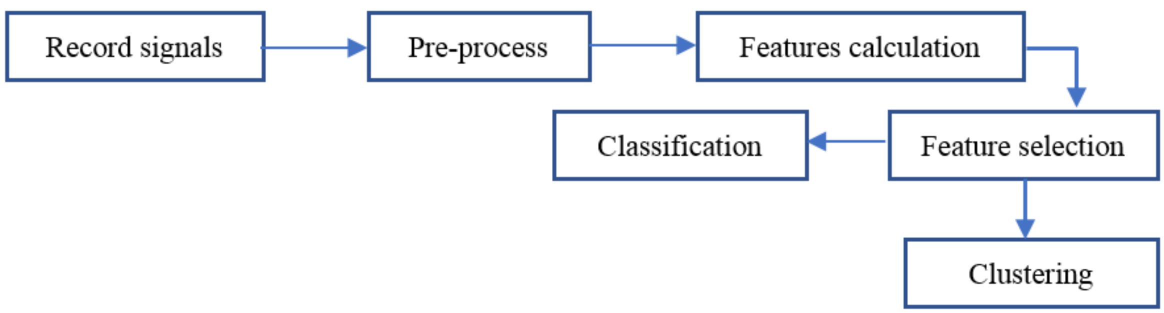


Figure 4. Physiological measure process

3.3.1 Signal recording

Three ANS signals - ECG, EDA, and RSP - were collected during both baselines and training scenarios. The signals were collected using the Nexus-10 MarkII data acquisition system with the accompanied Biotrace+ software (Mind Media Co., Herten, Netherlands). Sampling rates were 256 Hz for ECG and 32 Hz for EDA and RSP. Two

ECG electrodes were placed on the left and right chest, just below the clavicles, and one ECG electrode was placed just below the last rib on the left side. The two EDA electrodes were placed on the middle phalanxes of the middle and ring fingers. The RSP sensor band was worn around the participant's rib cage. See Figure 5 for sensor placement.

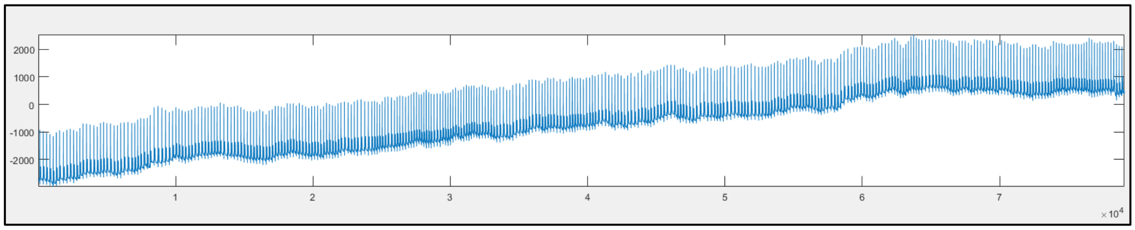
(a) (b)

Figure 5. Participant performing AVERT scenarios with physiological sensors placed on hands and chest. (a) electrodes 1, 2, 3 – ECG, sensor 4 – RSP belt (b) electrodes 5, 6 – EDA.

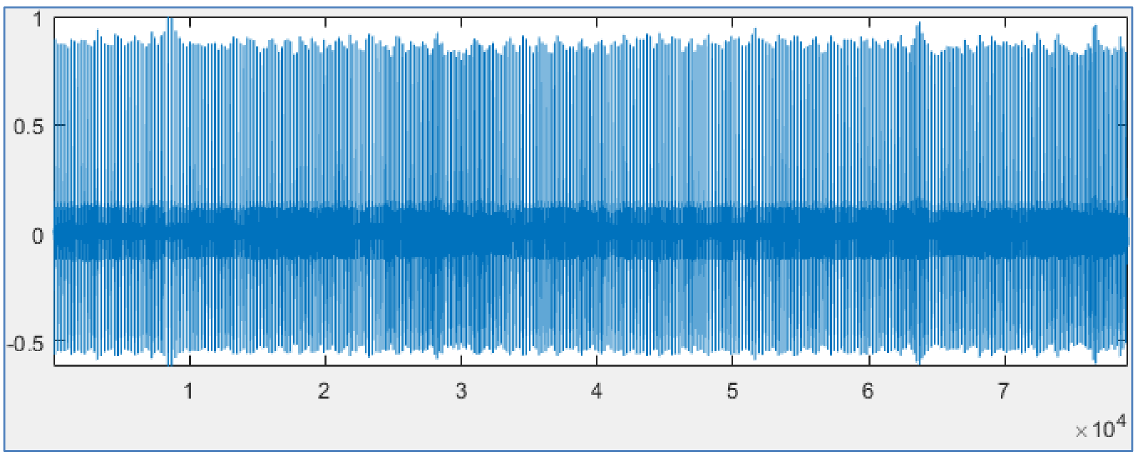
3.3.2 Signal pre-processing

The ECG, EDA, and RSP signals were first pre-processed to remove unwanted noise and to prepare them for feature extraction and classification:

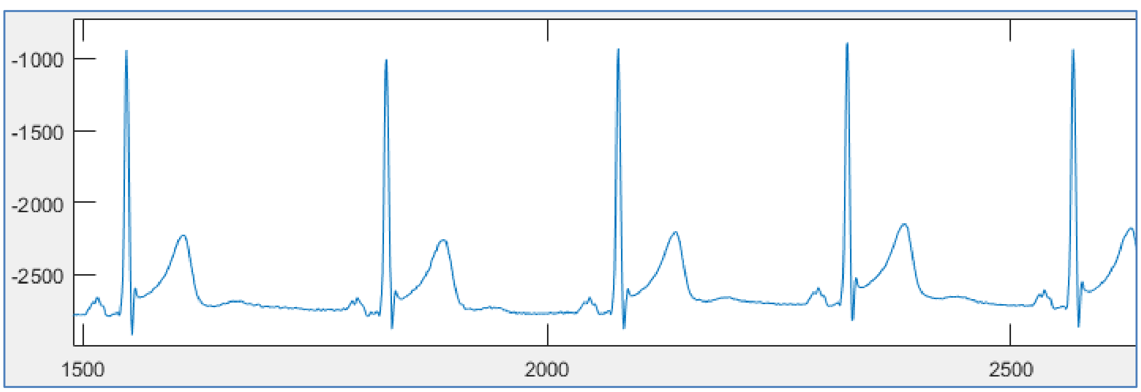
- The ECG signal was first filtered by a 5-15 Hz (Pan & Tompkins, 1985) 3rd-order Butterworth bandpass filter. As illustrated in Figure 6, the fluctuating raw ECG signal became flat (low frequency drift eliminated) after the bandpass filter was applied. The raw signal was not too noisy (did not contain unwanted high frequency components), and it was as clean as the filtered signal. This is because the recording device contained hardware filters to eliminate high frequency noise while recording.



(a)



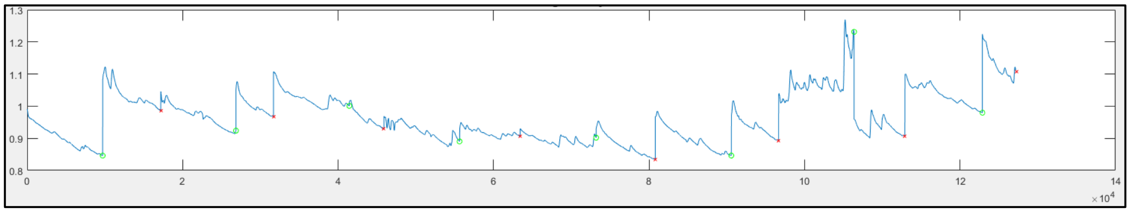
(b)



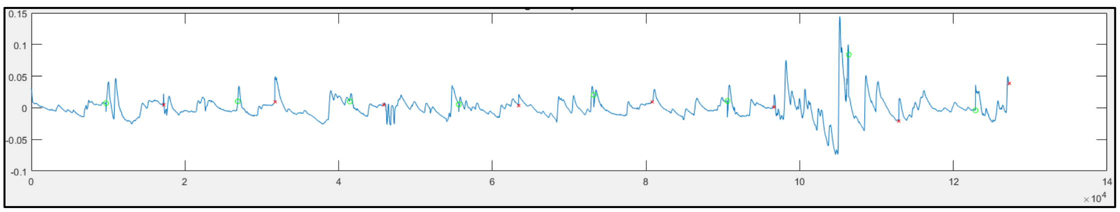
(c)

Figure 6. Illustrations of the effects of the bandpass filter on the raw ECG signal (a) raw ECG signal (b) bandpass filtered signal (c) a closer look at the raw ECG data

- The EDA signal was put through a 2nd-order Chebyshev lowpass filter with cut-off frequency of 1 Hz (Panju, et al., 2015), then detrended to eliminate possible linear trends. As can be seen from Figure 7, the filtered EDA signal was flatter, compared to the raw EDA signal. This shows that linear trends were eliminated from the raw signal.



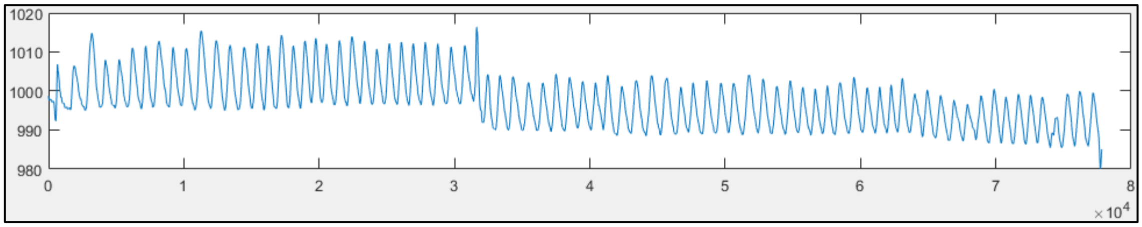
(a)



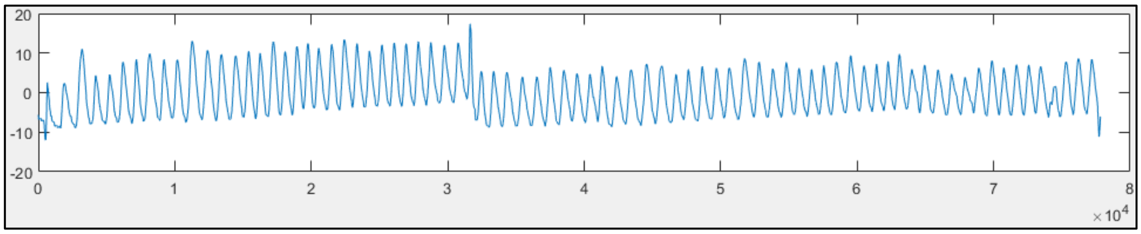
(b)

Figure 7. EDA signals (a) raw (b) pre-processed

- The RSP signal was detrended to eliminate any linear trend. As in the sample data illustrated in Figure 8(a), the raw RSP signal had a minor linear trend, which slightly decreased the signal over time, and this trend was eliminated as in Figure 8(b).



(a)



(b)

Figure 8. RSP signals (a) raw (b) pre-processed

3.3.3 Feature calculation

After pre-processing, all signals were segmented into 3-second intervals for feature extraction. All features were calculated from these 3-second segments. From the pre-processed ECG, a heart rate (HR) signal was calculated using the Pan-Tompkins algorithm (Pan & Tompkins, 1985). From this HR signal, seven different features of HRV were calculated (Malik, et al., 1996). From the pre-processed RSP signal, a respiration rate (RR) signal was calculated via a peak detection method developed by (Yoder, 2011). Calculation of these features will be described in more detail in the following section.

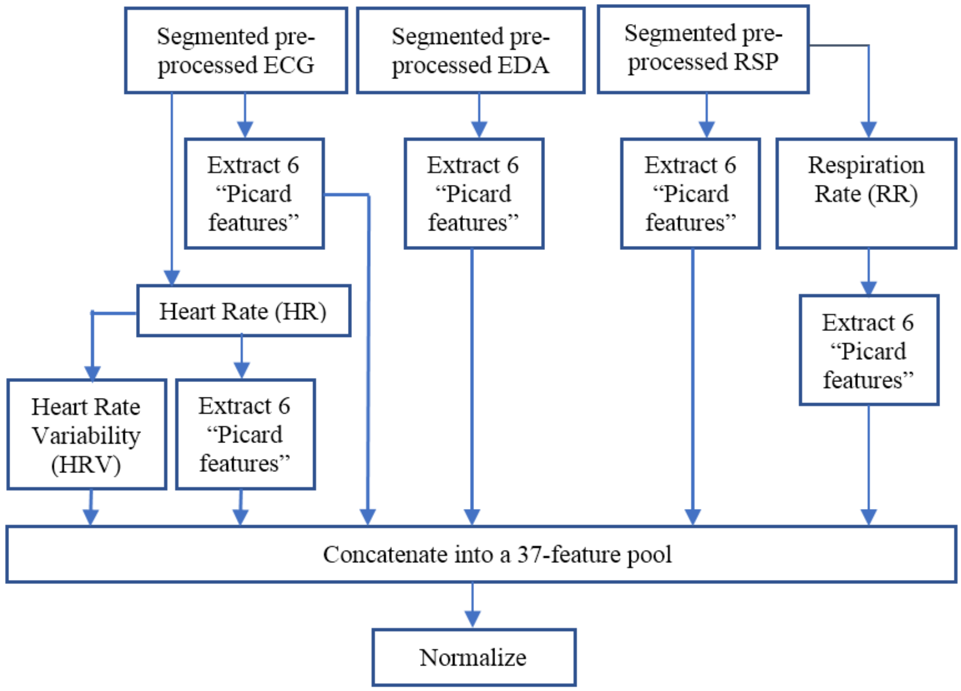


Figure 9. Features extraction process

The overall feature pool considered for classification consisted of the seven HRV measures, plus the following six characteristics calculated based on Picard et al.'s method (Picard, et al., 2001) for the pre-processed ECG, EDA, and RSP signals, as well as for the calculated HR and RR signals: 1) mean of the signal, 2) standard deviation of the signal, 3) mean of the absolute value of the first difference of the signal, 4) mean of the absolute value of the first difference of the normalized signal, 5) mean of the absolute value of the second difference of the signal, and 6) the mean of the absolute value of the second difference of the normalized signal. The resulting feature pool comprised 37 features. At this point, the data was normalized by Equation 1 to scale from 0 to 1, thus facilitating the process of classification because all dimensions now had values from 0 to 1, avoiding the case that the data was too skewed (some dimensions had a much larger or smaller scale than others). Figure 9 depicts the feature extraction process.

$$Normalized\ Data = \frac{Data - \min(Data)}{\max(Data) - \min(Data)} \tag{1}$$

3.3.3.1 QRS complexes detection in ECG signal

In order to calculate HR or HRV, QRS should be first detected from the ECG signal. QRS complexes were extracted from the preprocessed ECG signal through the following three steps (Pan & Tompkins, 1985):

Step 1: The filtered signal was differentiated to provide the QRS-complex slope information. A five-point derivative was used. The transfer function is:

$$H(z) = \frac{1}{8}(-z^{-2} - 2z^{-1} + 2z + z^{2}) \tag{2}$$

The frequency response of this derivative is nearly linear between 0 Hz (DC) and 30 Hz and its delay is 2 samples. A sample of ECG data following application of the derivative filter is shown in Figure 10(a).

Step 2: After differentiation, the signal was squared point-by-point to make all data points positive. This operation is non-linear and emphasizes the higher frequencies (i.e., predominantly the ECG frequencies). The equation of this operation is:

$$y[n] = x[n]^{2} \tag{3}$$

A sample of the data after point-by-point squaring is shown in Figure 10(b).

Step 3: Peak detection. The process of peak detection is described as follows:

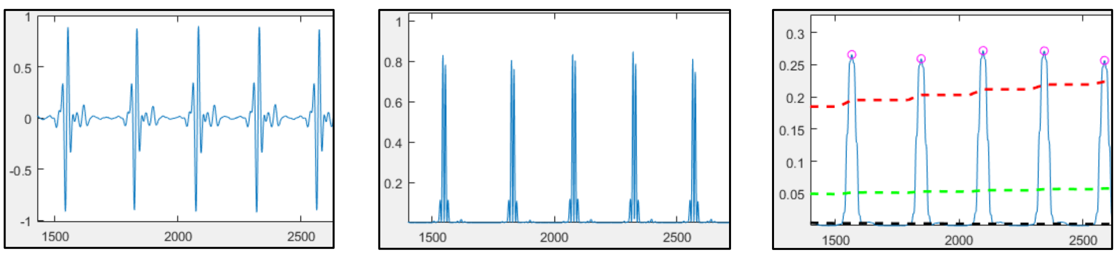
Firstly, the signal was integrated by a moving-window to obtain waveform feature information in addition to the slope of the R wave. It was calculated by Equation 4 below:

$$y[n] = \frac{1}{N}\left[x\big(n-(N-1)\big) + x\big(n-(N-2)\big) + \cdots + x(n)\right] \qquad (4)$$

where N is the width of the integration window in samples. It is important to choose an appropriate value of N. Generally, the width of the window should be approximately the same as the widest possible QRS complex. If it is too wide, the integration waveform will merge the QRS and T complexes together. In contrast, some QRS complexes will produce several peaks in the integration waveform. These can cause difficulty in subsequent QRS detection processes. The width of the window is determined empirically. In this case, the window width was chosen to be $0.15 \times$ (sampling rate), thus the window width was $0.15 \times 256 = 39$ samples.

After integrating the signal by the moving-window, a dynamic thresholding technique was applied to detect the peaks of the signal. Specifically, the algorithm used two threshold values (one for the true peaks and the other for the noisy peaks) that continuously adapt to changing ECG signal quality. After searching for the first time, the algorithm searches back for missed QRS complexes. At the first time, if a peak is lower than the signal threshold, it would not be considered an R-peak. However, if there is an unreasonably long period of time between two consecutive identified R-peaks, the algorithm will assume that an R-peak has been missed from the first scan. Therefore, at this time, the signal threshold and noise threshold are adjusted to capture those missing QRS complexes.

A sample of ECG data following application of this process is shown in Figure 10(c), where the circles represent the detected peaks, the upper and middle dashed lines represent the first and second signal thresholds, respectively, and the lower dashed line represents the noise threshold.



(a) (b) (c)

Figure 10. Illustration of ECG peak detection process: (a) derivative of signal (b) squared signal (c) detected peaks

3.3.3.2 Heart rate (HR) and heart rate variability (HRV)

Heart rate (HR)

HR is the information describing how fast a human's heart is beating, which is defined as thenumber of beats per minute. In this work, HR was calculated by the following equation:

$$HR = \frac{PP\ interval}{sampling\ rate} \times 60 \qquad (5)$$

where:

- PP interval is the peak-to-peak interval, the distance counted by number of samples between two consecutive R-peaks (QRS complexes)
- Sampling rate is 256 samples/second

Heart rate variability (HRV)

HRV is useful when investigating the changes in the electrical activity of the heart over time (Malik, et al., 1996). HRV can be calculated in many different ways, both in the time and frequency domains. Some of these measures require a long duration of ECG signal for accurate calculation. Due to the relatively short duration of signals in this work, seven different measures of HRV that could be accurately determined have been considered as features: 1) VLF (power in very low frequency range), 2) LF (power in low frequency range), 3) LF norm (LF power normalized), 4) HF (power in high frequency range), 5) HF norm (HF power normalized), 6) LF/HF, and 7) RMSSD (the square root of the mean of the sum of the squares of differences between adjacent normal-to-normal, or peak-to-peak, intervals).

3.3.3.3 Respiration rate (RR) calculation

Unlike ECG, which has a complex shape with different peaks in a single complex (Figure 11), the RSP signal has a simpler shape, which facilitates the process of peak detection. Firstly, raw RSP signal was read (Figure 11(a)) before a detrending filter was applied to remove possible linear trend in the signal (Figure 11(b)). Thereafter, a peak finding function was applied to reveal the peaks. Finally, RR was derived by the following equation (Figure 11(c)):

$$RR = \frac{PP\ interval}{sampling\ rate} \times 60 \qquad (6)$$

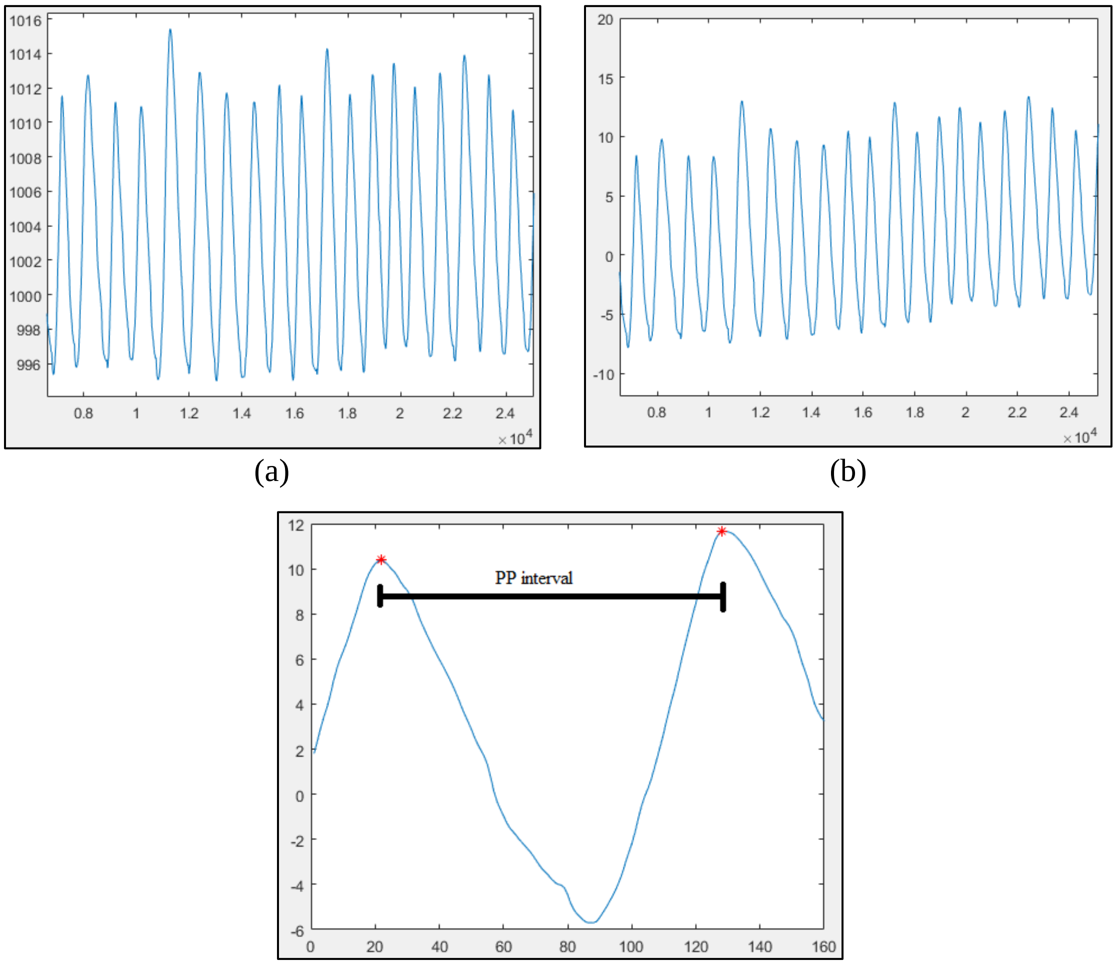


Figure 11. (a) Raw RSP signal (b) Detrended signal (c) Sample illustration of respiration peak-to-peak interval estimation

3.3.3.4 Picard's features

To recognize human emotional states, Picard, et al., (2001) proposed a set of six features, which were extracted from human physiological signals. As mentioned, in this work these six features, described below, were calculated for the pre-processed ECG, EDA, and RSP signals, as well as for the calculated HR and RR signals and included in the feature set for classification:

- The mean of the raw signal:

$$\mu_X = \frac{1}{N}\sum_{n=1}^{N} X_n \tag{7}$$

- The standard deviation of the raw signal:

$$\sigma_X = \left(\frac{1}{N-1}\sum_{n=1}^{N}(X_n - \mu_X)^2\right)^{1/2} \tag{8}$$

- The mean of the absolute values of the first differences of the raw signals:

$$\delta_X = \frac{1}{N-1}\sum_{n=1}^{N}|X_{n+1} - X_n| \tag{9}$$

- The means of the absolute values of the first differences of the raw signals:

$$\tilde{\delta}_X = \frac{1}{N-1}\sum_{n=1}^{N}\left|\tilde{X}_{n+1} - \tilde{X}_n\right| = \frac{\delta_X}{\sigma_X} \tag{10}$$

- The means of the absolute values of the second differences of the normalized signals:

$$\gamma_X = \frac{1}{N-2}\sum_{n=1}^{N-2}|X_{n+2} - X_n| \tag{11}$$

- The means of the absolute values of the second differences of the normalized signals:

$$\tilde{\gamma}_X = \frac{1}{N-2}\sum_{n=1}^{N-2}\left|\tilde{X}_{n+2} - \tilde{X}_n\right| = \frac{\gamma_X}{\sigma_X} \tag{12}$$

3.3.4 Feature selection

In order to automatically choose the best M-dimensional feature set from N-dimensional data, a feature selection algorithm can be used. A feature selection algorithm generally requires a search algorithm to efficiently search the feature space, and a fitness criterion by which to evaluate "how good" each candidate feature set within the space is. In this work, the Fisher score and a greedy search algorithm were used. Figure 12 depicts the feature selection process.

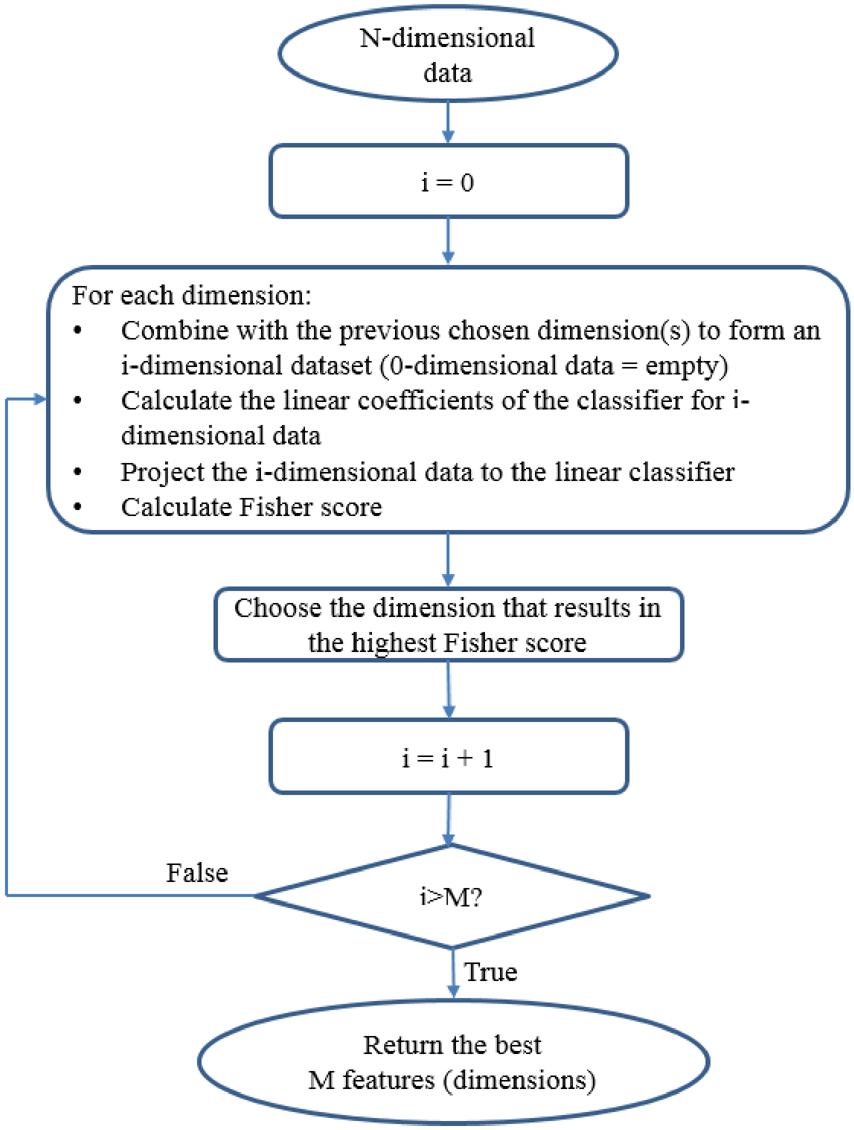


Figure 12. Feature selection process

3.3.4.1 Fisher criterion

The feature selection is based on the Fisher criterion (Gu, et al., 2011), which is widely used in classification to help find the dimension where data in two classes are separated the most. The Fisher score is calculated by the formula (13) below:

$$F(w) = \frac{|m_1 - m_2|}{s_1^2 + s_2^2} \tag{13}$$

where:

- m_1 is the mean of w^{th} dimension of the data (i.e., the w^{th} feature) for class 1
- m_2 is the mean of w^{th} dimension of the data for class 2
- s_1^2 is the variance of w^{th} dimension of the data for class 1
- s_2^2 is the variance of w^{th} dimension of the data for class 2

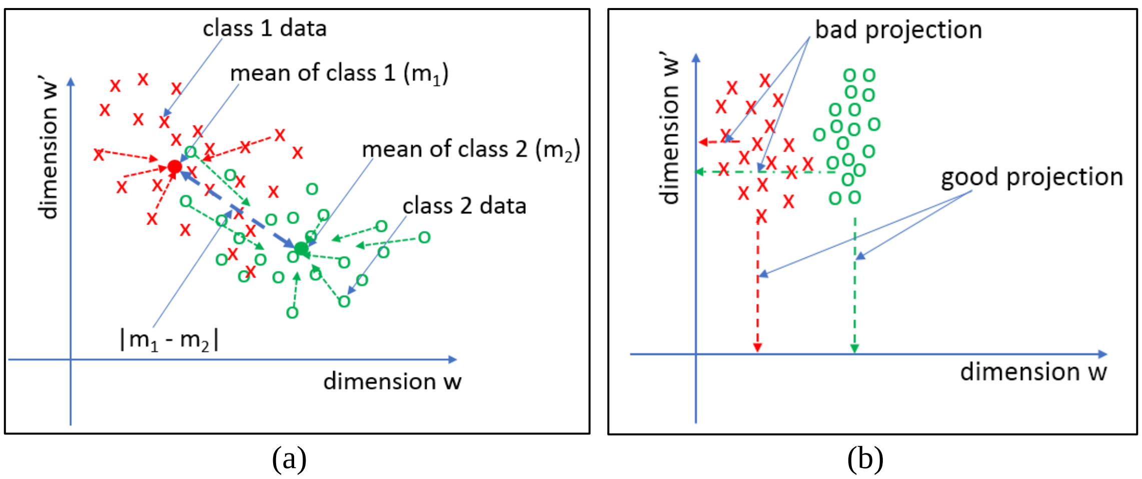


Figure 13. (a) Illustration of how the Fisher criterion works (b) Example of feature selection based on the Fisher criterion

From Figure 13(a), it can be seen that the Fisher criterion helps us choose the dimension that maximizes the distance between the mean of class 1 and 2, while minimizing the variance within one class. In other words, in the dimension recommended by the Fisher criterion, data in Class 1 and 2 are separated the most (maximum distance from two means), and data within a class are close to each other (minimum variance). This means data within a class do not spread out too much, thus reducing the chance that data from class 1 are mixed with data from class 2. Using this method, the best dimension could be picked out for the sake of the classification process. For example, in Figure 13(b), data in two classes 1 and 2 will be separated if they are projected to dimension w. Meanwhile, if projecting these data to dimension w', the data would be mixed together.

3.3.4.2 Greedy search procedure

Now that the dimension that has the highest Fisher score has been derived, another problem arising is that the best 3-feature set needs to be chosen, instead of only one best feature. An intuitive approach is the ranking system, where three features are selected that have the three highest Fisher scores. That sounds reasonable in one way, but the fact is not quite simple in another. Specifically, three features might have the highest Fisher scores when considered individually, but together they may not be the best possible combination of three features. There might be another set, constituted from other dimensions, that has a higher three-dimensional Fisher score. Consider the formula (13), where m_1 and m_2 are three-dimensional means, with m_1 is located at (x_1, y_1, z_1), and m_2 is located at (x_2, y_2, z_2), in a Cartesian coordinate system. The Fisher score is now calculated in a three-dimensional space. Therefore, the result is dependent on the three dimensions jointly, not independently. To achieve this, there is another intuitive approach, which is called “exhausted search”. This algorithm takes into account every possible combination of three features in the total number of features, calculating the Fisher score for each combination and selecting the one with the highest score. This method will result in the “best” possible feature set being selected, however is simply too costly in terms of computation when considering more than a modest number of candidate features. To compromise these two methods (ranking and exhaustive search), a greedy search method was chosen in this work. This method saves computational cost significantly, while still considering the joint aspect of the features. The greedy search process is divided into three steps, and these steps are replicated in a number of times equal to the number of dimensions we would like to select:

Step 1: Read the N-dimensional data

Step 2: Assign i = 0

Step 3: For each dimension:

- Combine with the previous chosen dimension(s) to form an i-dimensional dataset (0-dimensional data = empty)
- Calculate the linear coefficients of the classifier for i-dimensional data
- Project the i-dimensional data to the linear classifier
- Calculate Fisher score

Step 4: Choose the dimension that results in the highest Fisher score

Step 5: Increase i by 1

Step 6: Repeat the actions from step 3 until i > M

Step 7: Return the best M features (dimensions)

3.3.5 Classification

Two classifiers were employed, linear discriminant analysis (LDA) and support vector machine (SVM), and the results were compared to see which method was better to solve the problem. Cross validation was applied to find the classification accuracy.

3.3.5.1 Discriminant Analysis Classifiers

After picking out the best feature set by the Fisher criterion, classifiers based on discriminant analysis were used to classify baseline and scenario data. In this method, data in each class are assumed to have a Gaussian mixture distribution. Weighted classifiers are constructed using a scheme described in (Fisher, 1936). The result is a linear or quadratic discriminant analysis (LDA or QDA) classifier (Figure 14).

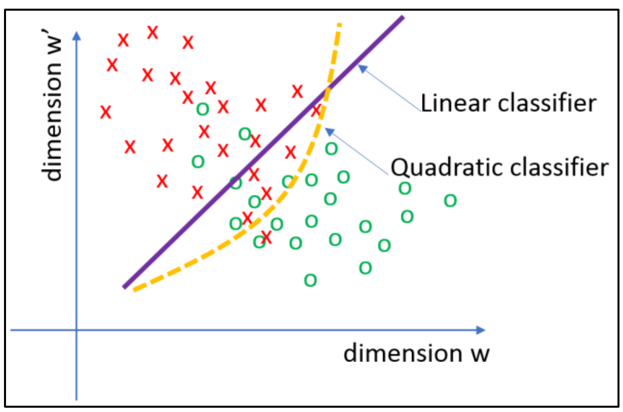
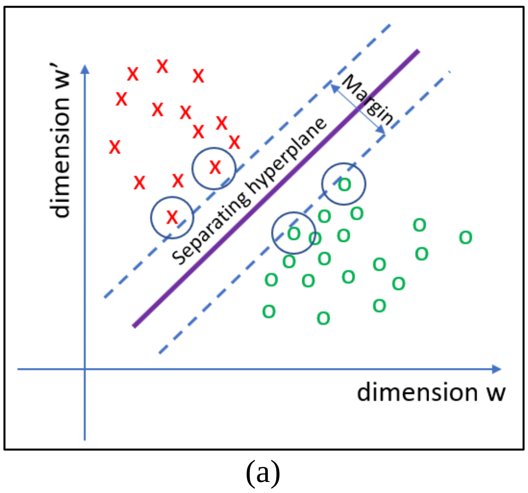


Figure 14. Illustration of linear and quadratic discriminant analysis classifiers

3.3.5.2 Support vector machine (SVM)

An SVM algorithm (Kecman, 2001; Suykens, et al., 2002; Scholkopf & Smola, 2002; Cristianini & Shawe-Taylor, 2000) with Gaussian kernel was also employed to classify the data between each baseline and scenario for every participant individually. By doing this, possible implementation errors could be avoided because results from two different classifiers are derived and compared with each other. SVM is a widely-used machine learning algorithm, which offers a reserved space for the model, thus preventing the model from overfitting the trained data in the case that data in different classes are separable (Figure 15).



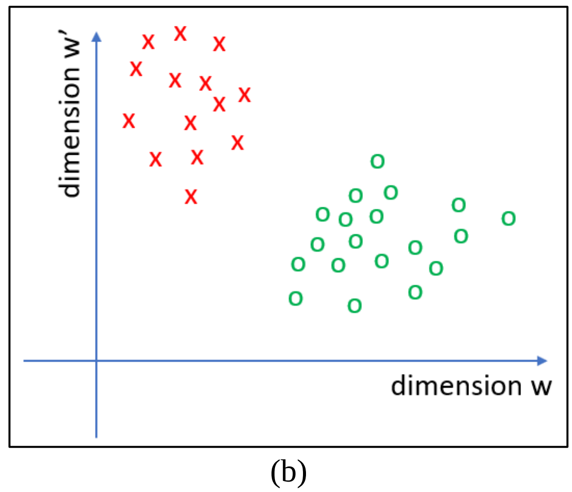


Figure 15. (a) Illustration of SVM classifier (b) an example of consequences of not having reserved space for classifier when data in different classes are separable

3.3.5.3 Cross validation

In this work, 30 runs of 5-group cross validation (Christopher, 2016) were performed for each baseline versus scenario condition for each of the classification algorithms considered. Data in each class (baseline and scenario) is presented as a $[M \times N]$ matrix, where M is the number of data points (i.e., the number of 3-second intervals) and N is the number of features considered. These data points were randomly divided into five groups, with roughly equal numbers of each class in each group (the odd number of data were eliminated in case the total number of data was not a multiple of five). In the first group of the cross-validation, four of the groups were combined to be training data, while one group was left out to be testing data. Feature selection was performed on the training data and a classifier was built and then tested on the test data to determine classification accuracy for this first iteration. In the next group of the cross-validation, a different group of data was left out to be used for testing, while the remaining four were combined to make up the training data. Classification accuracy was again determined for this iteration. This was repeated a total of five times to create five different training data sets and five corresponding testing data sets, such that each group of data was used as the test set one time. Following the five groups of the cross-validation, the mean classification accuracy was calculated. 30 runs of this 5-group cross-validation procedure was completed (with the data being randomly divided into groups at the beginning of each run), and the overall accuracy was calculated as the mean of the 30 runs. This process is illustrated in Figure 16.

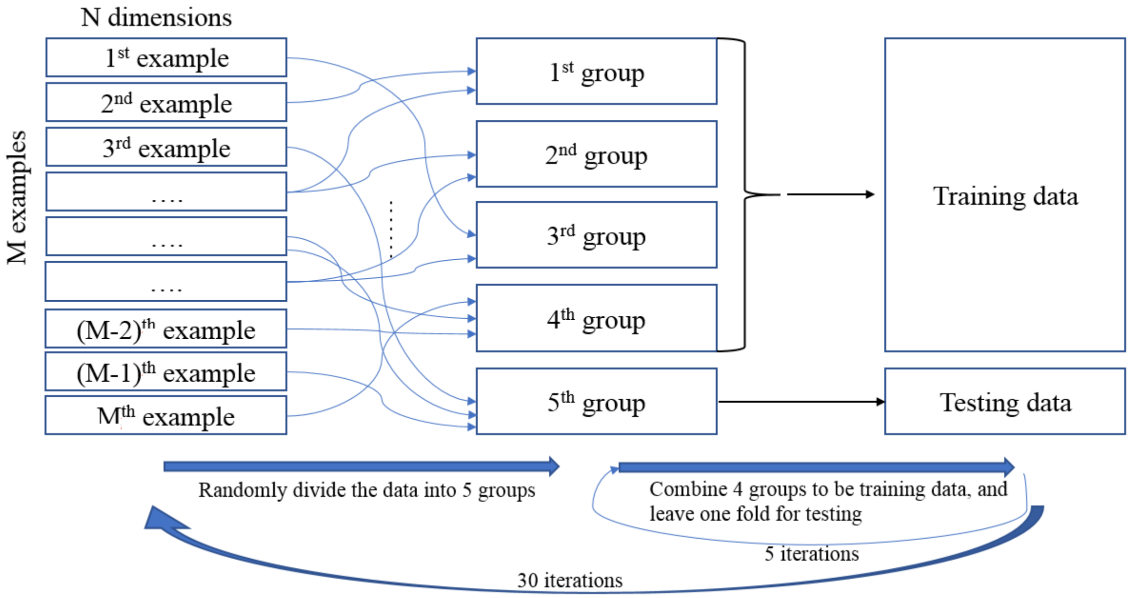


Figure 16. Cross-validation process

Note that normally, when data from two classes are balanced, or have the same number of samples, accuracy is calculated by the formula below:

$$Accuracy\ (\%) = \frac{TP + TN}{Total} \times 100 \tag{14}$$

where:

- TP – True positive: number of testing data in class 2 (scenario) correctly classified as class 2
- TN – True negative: number of testing data in class 1 (baseline) correctly classified as class 1
- Total: total number of testing data points (baseline + scenario)

However, in this work, data from the baseline were not at the same length as the ones from the scenario. Therefore, adjusted accuracy was calculated from sensitivity and specificity to make up for the skewness of the data (Zeng, et al., 2002). Besides TP and TN denoted, let us denote FN as false negative – number of scenario data points incorrectly classified as baseline, and FP as false positive – number of baseline data points incorrectly classified as scenario. From that, sensitivity, specificity, and adjusted accuracy were defined as:

$$Sensitivity\ (\%) = \frac{TP}{TP + FN} \times 100 \tag{15}$$

$$Specificity\ (\%) = \frac{TN}{TN + FP} \times 100 \tag{16}$$

$$Adjusted\ accuracy\ (\%) = \frac{Sensitivity + Specificity}{2} \tag{17}$$

3.3.6 Clustering

Along with supervised machine learning (discriminant analysis and SVM), unsupervised machine learning algorithms were also implemented to explore possible patterns in the data. In this method, principle component analysis was first applied to reduce the dimensionality of the data, then two unsupervised learning algorithms were employed to extract possible patterns. The algorithms included k-means clustering and Gaussian model clustering.

3.3.6.1 Principle component analysis (PCA)

In unsupervised machine learning, it is also necessary to reduce the number of dimensions of the data before conducting the classification process, in order to save computational cost, and also because not every feature is useful. Unlike the supervised learning case, where we could use the Fisher criterion to select the feature combination that best separates data in two classes because we know which data points belongs to which class, we could not do the same for the unsupervised learning case because the information of classes is unknown. In this case, PCA is a possible solution. This method reduces the data dimensionality by projecting data into new dimensions – called principled components – that are calculated from the existing features (Figure 17). Generally, most of the information in the data is contained in a relatively small number of these new features. Thus the desired number of features can be retained and the rest discarded, resulting in a reduction in the dimensionality of the data.

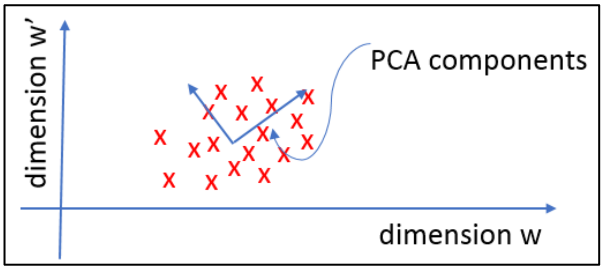


Figure 17. PCA components

3.3.6.2 K-means clustering

After a number of clusters (k) were assigned, the algorithm started grouping data to k groups. In order to achieve reasonable grouping results, k centroids were first initialized, then the Euclidian distance from each data point to the centroids was calculated. Each data point was assigned to be in the group where the distance from the data point to the centroid was smallest compared to the distances from the data point to the other centroids. After that, the total distance of all data points to their centroids were calculated, and by the optimization process, new centroids were created which minimized the total distance. The process was iterated until new centroids were not significantly different from the previous one. Illustrations for the process are provided in Figure 18.

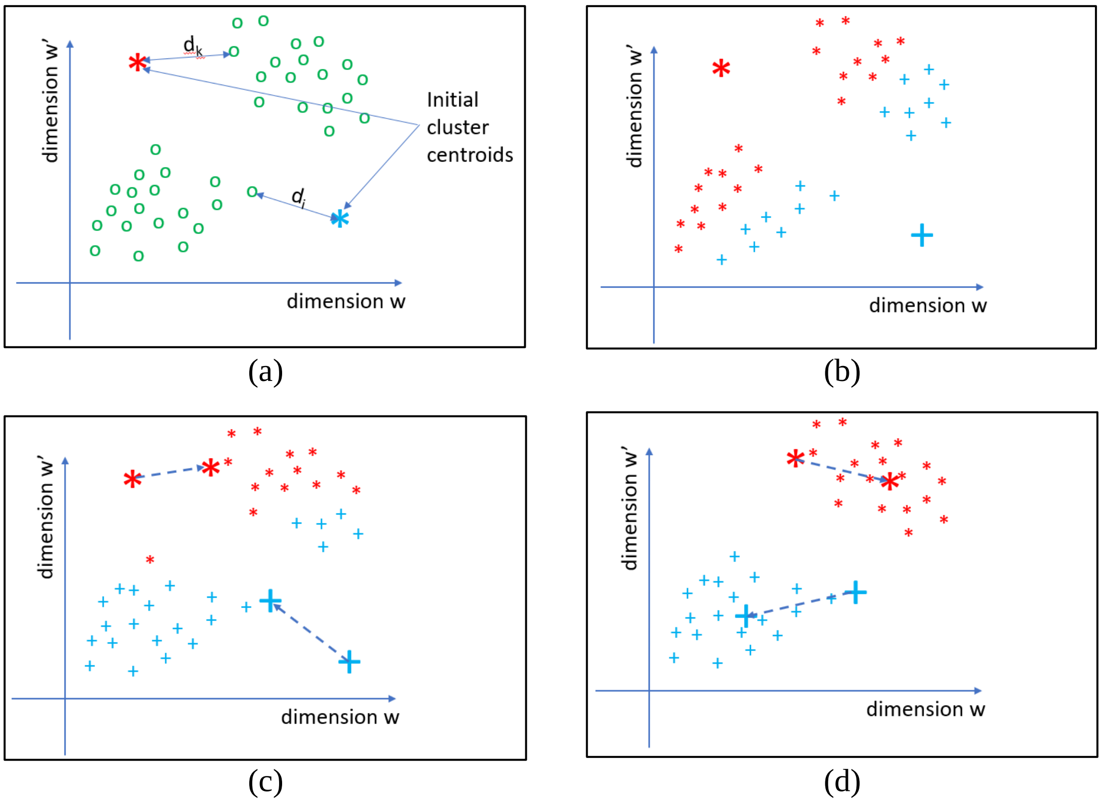


Figure 18. K-means clustering process (a) randomly create initial centroids (b) assign data points into each group (c) move the centroids to minimize the total distance (d) iterate until reaching optimal point

3.3.6.3 Gaussian model clustering

While the k-means clustering method groups data based on optimizing the total distance from data to centroids, the Gaussian model-based method focuses on fitting data into multivariate Gaussian distribution. This method does not require information about the number of clusters, k. Instead, it tries to find the Gaussian distribution component(s) that best fit the given data (McLachlan & Peel, 2004). This method is capable of not only grouping data, but also detecting anomaly data points (Figure 19).

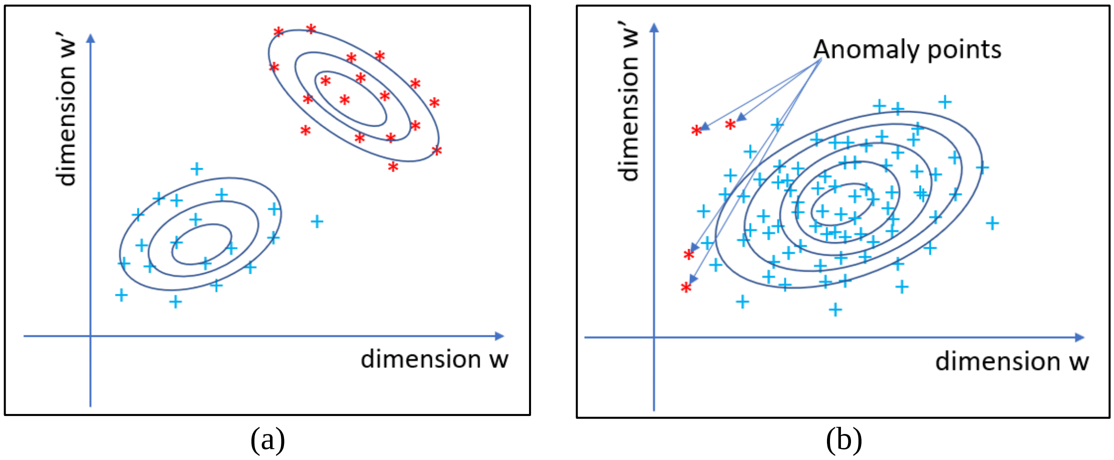


Figure 19.Gaussian model-based method's applications (a) grouping data (b) anomaly detection

3.4 Questionnaires

The participant's subjective rating of stress experienced during the performance of each scenario was recorded. In particular, following each trial, participants were asked a question, which was specifically made to fulfill the purpose of this research. The question was: "How did you feel during the scenario you just completed?" and the participants were asked to provide a rating from 1 (very relaxed) to 7 (very stressed).

3.5 Statistical analysis

After the three measures of interest were calculated (performance score, subjective stress rating, and classification accuracy), a factor analysis was conducted to explore which of the three performance shaping factors, or interactions between them, had significant effects on the responses. Additionally, a repeated measures ANOVA was implemented to investigate whether there was a significant difference in responses among scenarios, which would indicate that different scenarios led to different experiences for trainees.

The order in which the eight scenarios were completed was randomized for each participant (Table 4). In this table, each row represents the sequence of scenarios taken by a participant. For example, participant number 1 went through the process of eight scenarios in the order of: scenario 7, scenario 8, scenario 6, scenario 1, scenario 3, scenario 4, scenario 2, and scenario 5. Note that in this work, the term "scenario 1" represents the name of a specific scenario with a specific combination of PSFs, while "1st scenario" represents the first scenario in terms of the time order that a participant experienced.

Table 4. The sequence of scenarios taken by each participant

Subject	1st scenario	2nd scenario	3rd scenario	4th scenario	5th scenario	6th scenario	7th scenario	8th scenario
1	7	8	6	1	3	4	2	5
2	5	4	1	8	7	3	2	6
3	6	2	8	7	3	5	1	4
4	8	2	4	7	1	6	3	5
5	3	7	8	2	1	5	6	4
6	7	2	5	3	1	8	6	4
7	4	6	7	3	1	5	2	8
8	5	3	2	4	6	7	8	1
9	6	7	5	3	1	2	4	8
10	3	6	2	7	5	1	4	8
11	5	8	7	4	1	2	6	3
12	3	5	1	4	2	8	7	6
13	4	8	2	5	6	7	1	3
14	5	3	6	7	4	2	8	1
15	4	8	3	1	5	7	6	2
16	7	5	4	6	1	2	3	8
17	4	1	5	7	3	6	2	8
18	2	6	3	5	4	7	8	1
19	7	5	8	6	4	2	3	1
20	6	1	8	7	4	2	3	5
21	2	5	3	4	1	6	7	8

22	2	5	6	7	4	8	1	3
23	7	6	2	8	3	5	1	4
24	7	3	8	6	2	5	4	1
25	8	1	2	7	3	5	6	4
26	1	5	6	4	7	8	3	2
27	7	8	6	3	4	2	5	1
28	5	4	7	8	1	3	6	2
29	2	1	4	6	7	8	5	3
30	3	1	4	5	2	6	7	8
31	4	7	8	1	5	6	3	2
32	4	2	1	3	6	5	8	7
33	5	4	2	1	8	6	7	3
34	6	8	7	2	5	3	4	1
35	3	6	8	5	4	7	1	2
36	2	8	1	6	4	7	5	3
37	6	8	4	1	5	7	2	3
38	4	1	6	7	2	8	3	5

Considering the case that physiological arousal due to the scenarios could potentially decrease over the course of the experiment, for example as participants simply become used to the virtual environment and the being in an experimental setting, a linear regression was conducted to see whether there was a time trend for physiological changes in this experiment. The classification accuracy values from order based on scenario name (scenario 1, 2, 3, etc.) were re-arranged to order based on time (from the first scenario that a participant completed to the last one, in a chronological order).

Finally, a correlation analysis was employed to find any relationships among the subjects' performance scores, physiological arousal levels (as indicated by classification accuracy between baselines and scenarios), and subjective measures of stress.

Chapter 4: Results

4.1 Performance measures

All scores calculated from the rubrics were scaled to 100% as shown in Figure 20[3]. The results are presented in the format of Box-Whisker plot. For example, in scenario 4, the minimum and maximum performance of 37 participants are 81% and 100%, respectively. There is one participant considered as an outlier, whose performance is about 17%. The first quartile is approximately 87%, and in this case the third quartile is equal to the median value, which is 95%. Recall that there were 8 different scenarios, and 38 subjects completed the experiment. As can be seen in Figure 20, although there are a small number of outliers in some scenarios that have low performance (<40%), the data shows that participants performed well (>80%) most of the time.

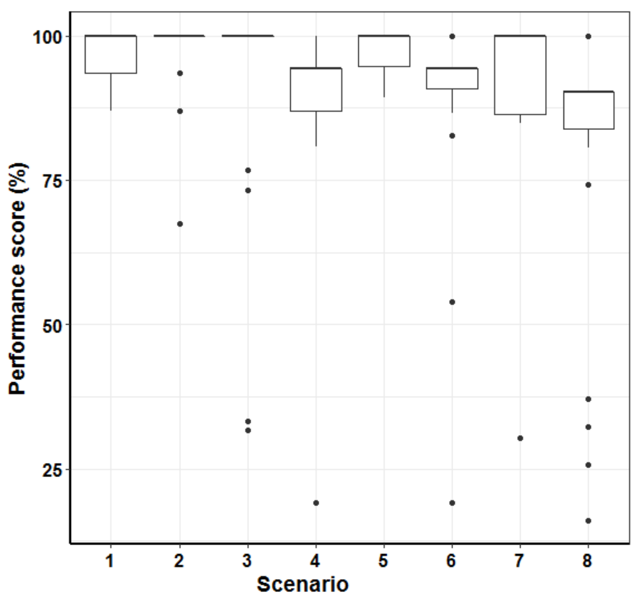


Figure 20. Performance score Box plot

[3] Detailed results are provided in Table 19 - Table 26

4.2 Physiological measures

4.2.1 Classification

The adjusted accuracy calculated from the 3-feature LDA algorithm is provided in Figure 21(a)[1]. It can be seen that the average classification in each scenario is relatively high (between 78.3% and 83.0%, with mean 81.3% across scenarios). The results from the 3-feature SVM algorithm are presented in Figure 21(b)[1]. The average classification accuracies derived from this algorithm are generally lower than those from the 3-feature LDA (between 68.8% and 78.8%, with mean 74.6% across scenarios). In the case of utilizing full features, the results from LDA and SVM algorithms are presented in Figure 21(c) and Figure 21(d), respectively[4]. There was an increase in classification of full-feature LDA and SVM compared to the 3-feature ones. Specifically, the mean of accuracies derived by full-feature LDA ranged from 81.5% to 85.8%, with a mean of 84.0% across scenarios, and for full-feature SVM mean accuracies ranged from 75.7% and 82.4% with a mean of 79.4% across scenarios.

[4] Detailed results are provided in Table 27 - Table 30

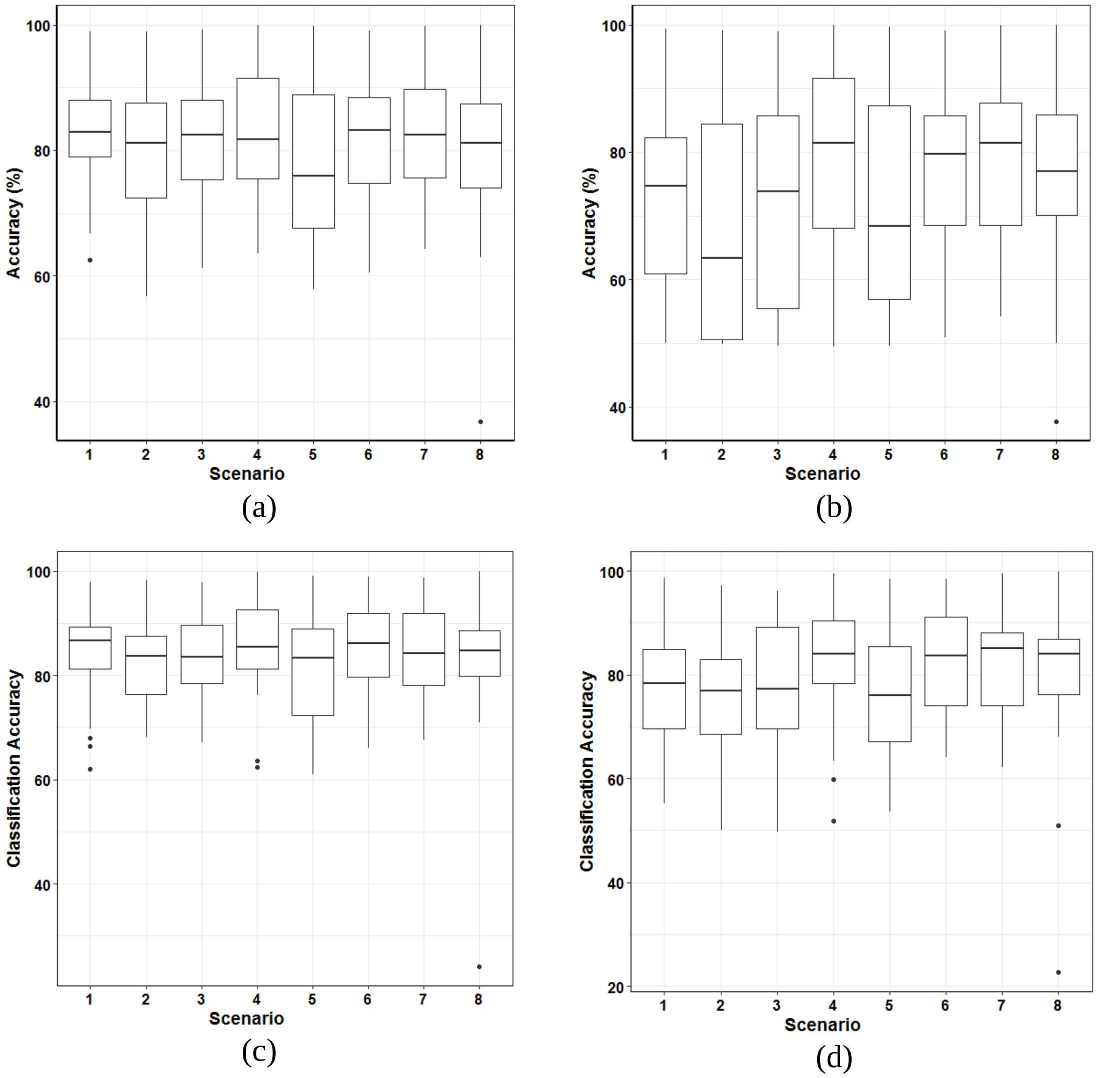


Figure 21. Summary of classification results from (a) 3-feature LDA (b) 3-feature SVM (c) full-feature LDA (d) full-feature SVM

4.2.2 Clustering

In clustering, participant data were analyzed both individually and combined. Specifically, the clustering algorithms were implemented on four forms of data:

- Cluster all data (scenario and baseline combined) of each participant
- Cluster all scenario data (excluding baseline data) of each participant

- Cluster all data (scenario and baseline combined) of all participants combined
- Cluster all scenario data (excluding baseline data) of all participants combined

Before clustering algorithms were applied, PCA was implemented to reduce the dimensions of the data to the number of dimensions where at least 99% of the information was retained. The number of PCA components chosen for each case is presented in Table 5:

Table 5. Different cases of applying clustering

Case	Number of PCA components
Cluster all data (scenario and baseline combined) of each participant	9
Cluster all scenario data (excluding baseline data) of each participant	9
Cluster all data (scenario and baseline combined) of all participants combined	20
Cluster all scenario data (excluding baseline data) of all participants combined	20

4.2.2.1 K-means clustering

K-means clustering was conducted with K=4 and results from k-means clustering are presented in the form of the percentage of each scenario data that fell into a cluster. The author expected to see a pattern where some scenarios might have most of their data belonging to some clusters while other scenarios have most of their data belonging to the other clusters.

- *Case 1: Cluster all data (scenario and baseline combined) of each participant*

In this case, physiological data collected from each participant through all sessions (baselines and scenarios) were combined and all labels were removed (unsupervised learning) before the algorithm was applied to separate data into four clusters. Thereafter, the number of data points from each scenario and baseline that belong to each cluster was counted and transferred into proportion (out of 100%), to see if there is any possible pattern. The outcomes are presented in Figure 22[5] (note that scenario 0 represents baseline).

[5] Detailed results are provided in Table 31 - Table 38

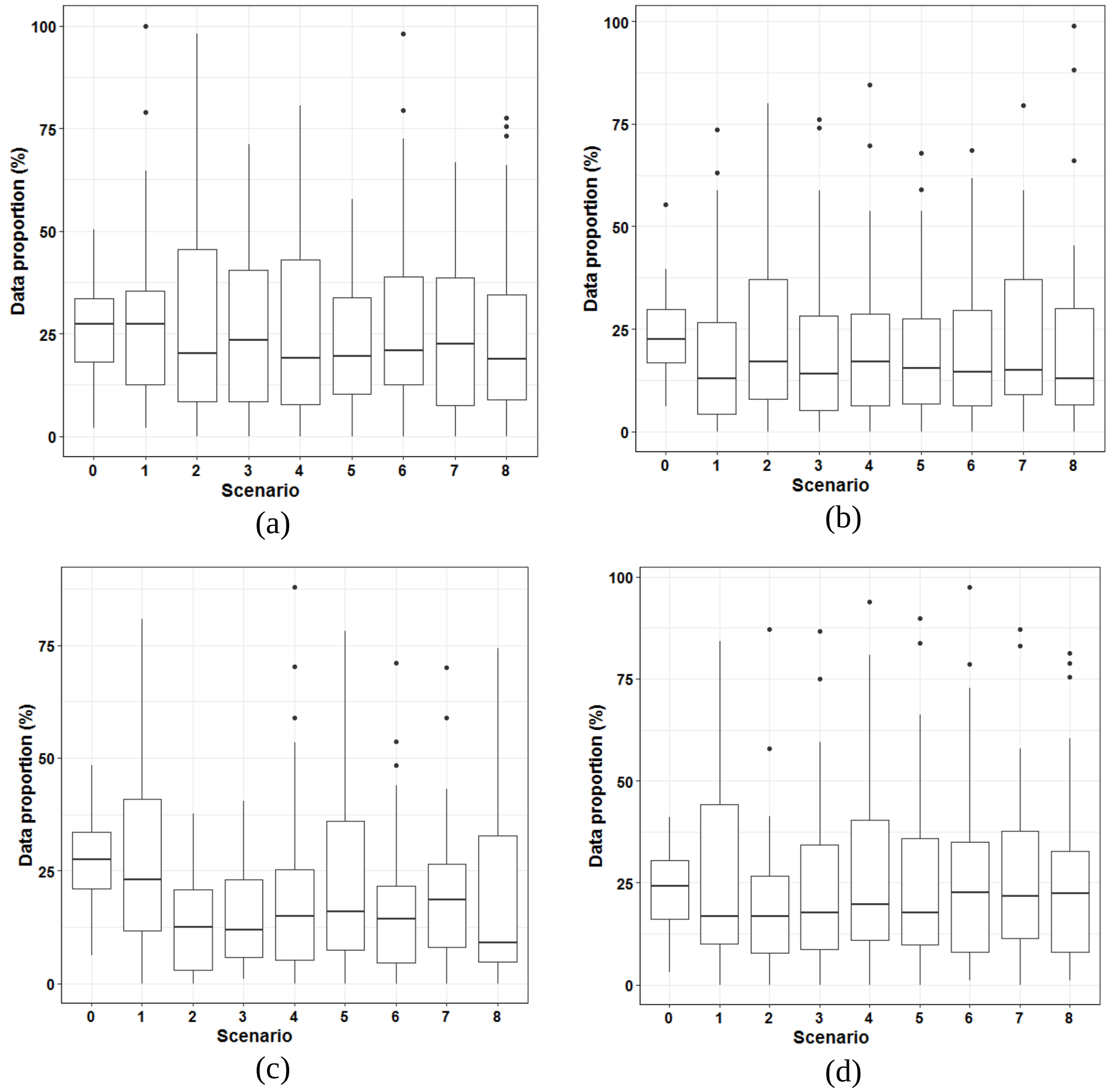


Figure 22. Summary of data proportion of each participant in (a) cluster 1 (b) cluster 2 (c) cluster 3 (d) cluster 4

- *Case 2: Cluster all scenario data (excluding baseline data) of each participant*

The process of this case is the same as the first case, except baseline data were excluded. The results are illustrated in Figure 23[6].

[6] Detailed results are provided in Table 39 - Table 46

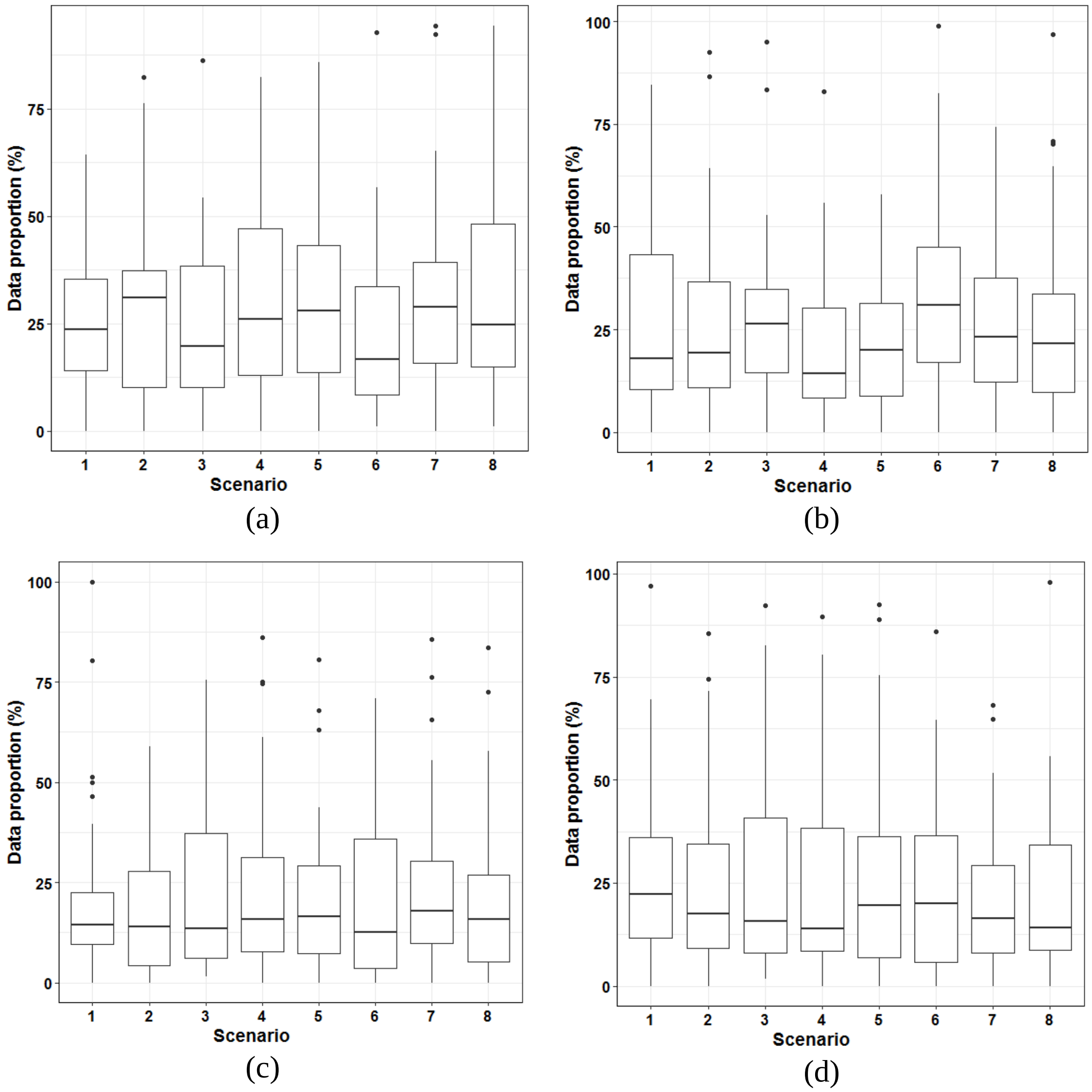
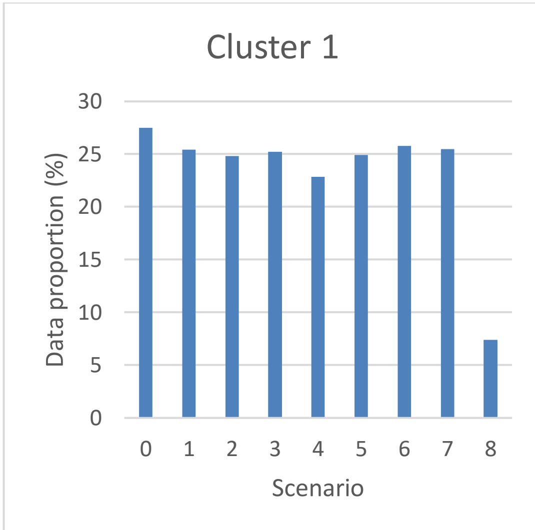


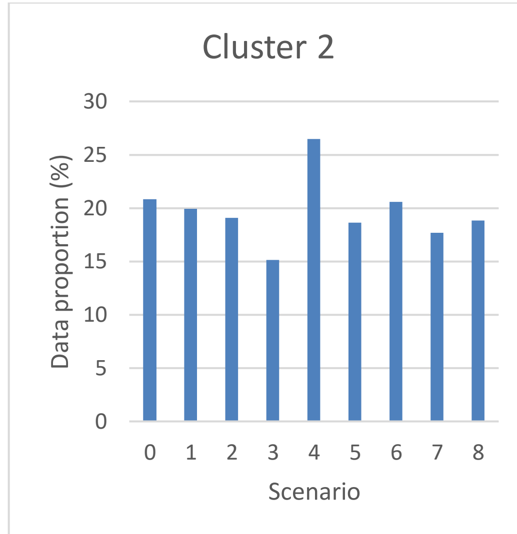
Figure 23. Summary of data proportion of each participant in (a) cluster 1 (b) cluster 2 (c) cluster 3 (d) cluster 4

- *Case 3: Cluster all data (scenario and baseline combined) of all participants combined*

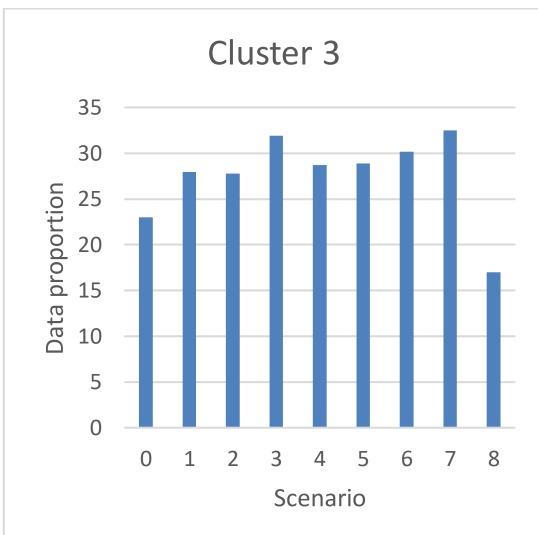
In this case, all data, including scenario and baseline data collected from each participant, were combined together before the clustering algorithm was applied to divide them into four different groups. The results are depicted in Figure 24 (note that scenario 0 represents baseline). The results in this case are presented in histogram form, instead of Box-Whisker plot. The reason is that the result for each scenario from all data from 38 participants are now combined instead of individual as in previous cases. Then, there is only one output value for each scenario, instead of 38.



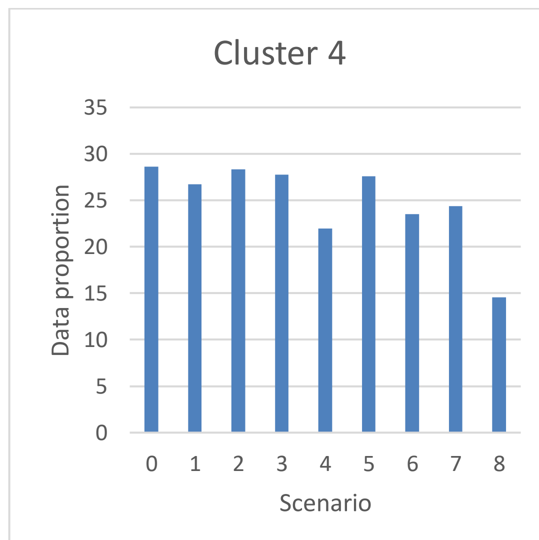
(a)



(b)



(c)



(d)

Figure 24. Summary of data proportion of each participant in (a) cluster 1 (b) cluster 2 (c) cluster 3 (d) cluster 4

- *Case 4: Cluster all scenario data (excluding baseline data) of all participants combined*

The process used in this case is the same as case 3, except baseline data were excluded. The results are shown in Figure 25.

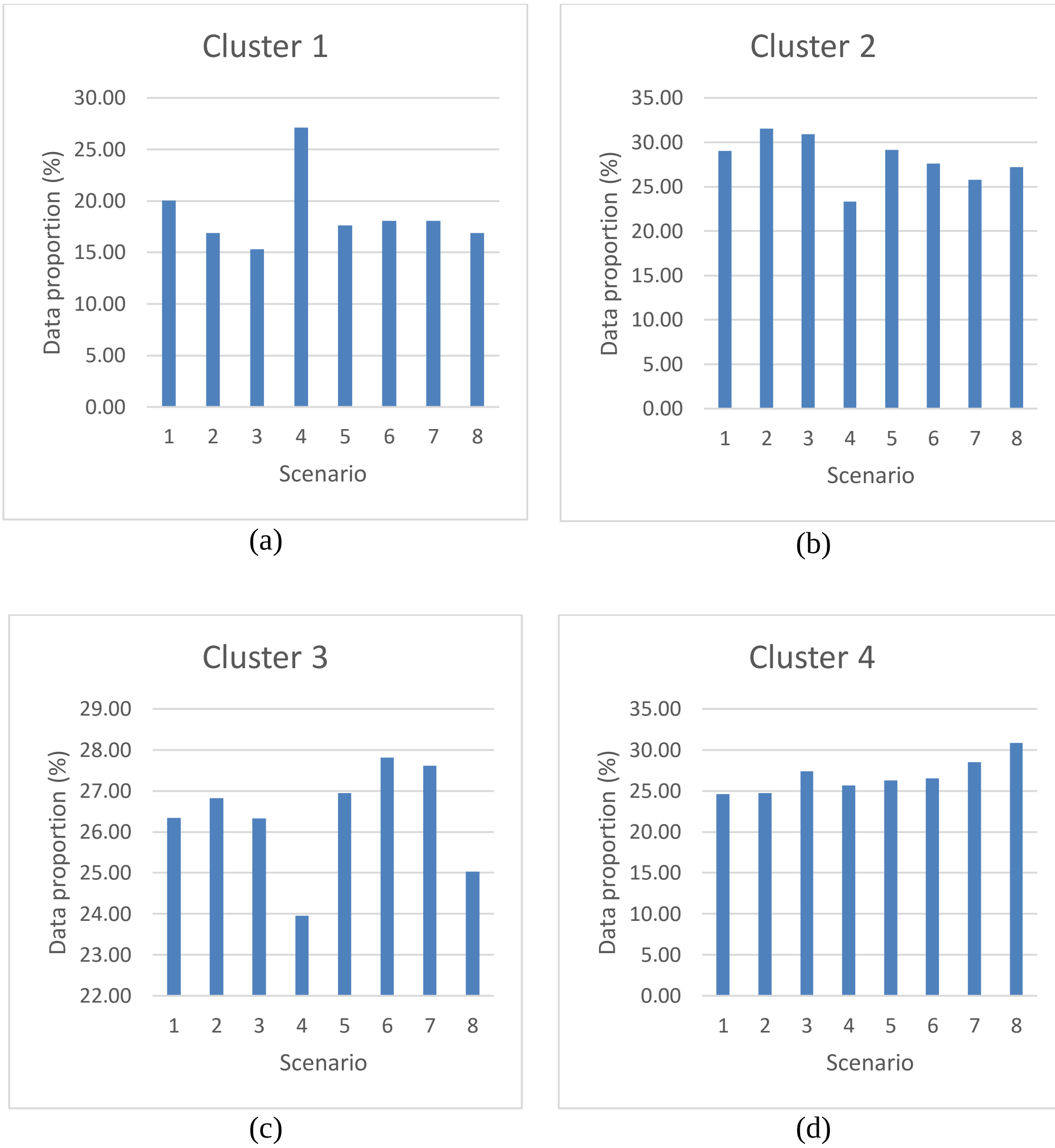


Figure 25. Summary of data proportion of each participant in (a) cluster 1 (b) cluster 2 (c) cluster 3 (d) cluster 4

4.2.2.2 Gaussian model clustering

In addition to k-means clustering, Gaussian model clustering was also applied to find a possible pattern in the data. Unlike k-means clustering, which was applied to divide data into k (in this case, four) clusters, Gaussian model clustering was implemented to find anomaly points and count them. The results reported are the number of anomaly points detected in each scenario in four cases (Table 5).

- *Case 1: Cluster all data (scenario and baseline combined) of each participant*

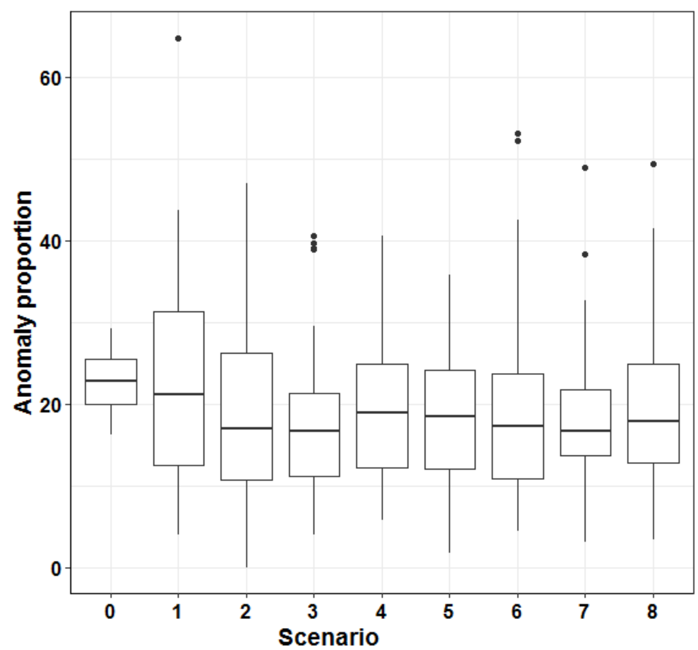


Figure 26. Proportion of anomaly points detected from multivariate Gaussian distribution model

In this case, a multivariate Gaussian distribution model of physiological data collected from each participant (including both scenario and baseline) was built and anomaly points were detected by a threshold of 10%. This means that any data point that fell into the region where the probability of data appearing was less than 10% and was recognized as an anomaly point. The number of anomaly points was then counted and tracked back to the origin to see which scenario or baseline it belonged to. These numbers were then transferred into proportion to reflect how much of each scenario or baseline data fell out of the majority. The results are provided in Figure 26[7] (note that scenario 0 represents baseline).

- *Case 2: Cluster all scenario data (excluding baseline data) of each participant*

This is the same as case 1, except the baseline data were excluded. The results are shown in Figure 27[8].

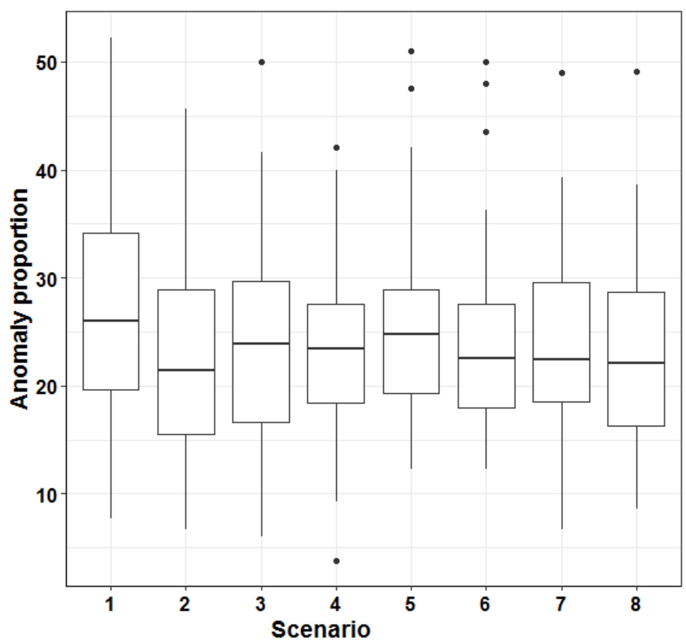


Figure 27. Proportion of anomaly points detected from multivariate Gaussian distribution model

[7] Detailed results are provided in Table 47 and Table 48
[8] Detailed results are provided in Table 49 and Table 50

- *Case 3: Cluster all data (scenario and baseline combined) of all participants combined*

In this case, all data, including scenario and baseline data of all participants, were combined and a multivariate Gaussian distribution of this data set was estimated. Thereafter, anomaly points were detected by the threshold of 1% and tracked back to see which scenario or baseline they belonged to. It is noticed that the threshold in this case is 1%, which is different than the other cases (10%). This is because when combining all data, the number of data points was very large and spreading, leading to the fact that the anomaly proportion was quite similar for every scenario. Therefore, reducing the threshold would provide more insight into the difference among scenarios in terms of anomaly proportion. These numbers were then transferred into proportion. The outcomes are illustrated in Figure 28.

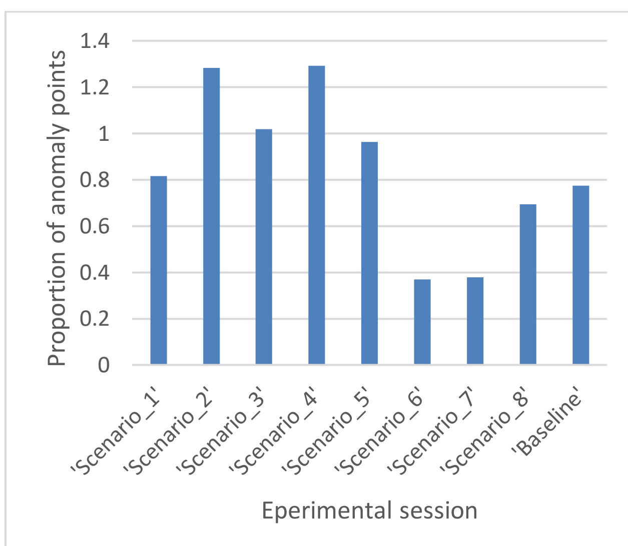


Figure 28. Proportion of anomaly points detected from multivariate Gaussian distribution model

- *Case 4: Cluster all scenario data (excluding baseline data) of all participants combined*

The process of this case is the same as case 3, except baseline data were excluded. The results are shown in Figure 29.

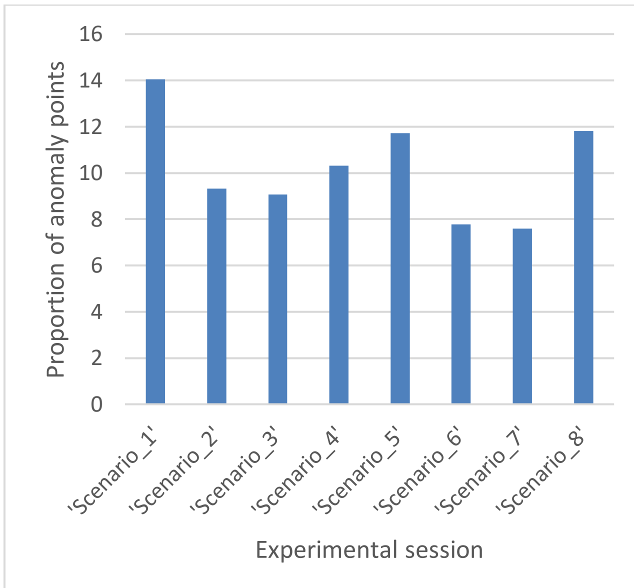


Figure 29. Proportion of anomaly points detected from multivariate Gaussian distribution model

4.3 Subjective ratings of stress

Results for subjective ratings of stress are provided in Figure 30[9], note that the rating scale was from one to seven. The most frequent responses for the questionnaires range from 1 to 4, which indicate low stressful levels during the experiment.

[9] Detailed results are provided in Table 51

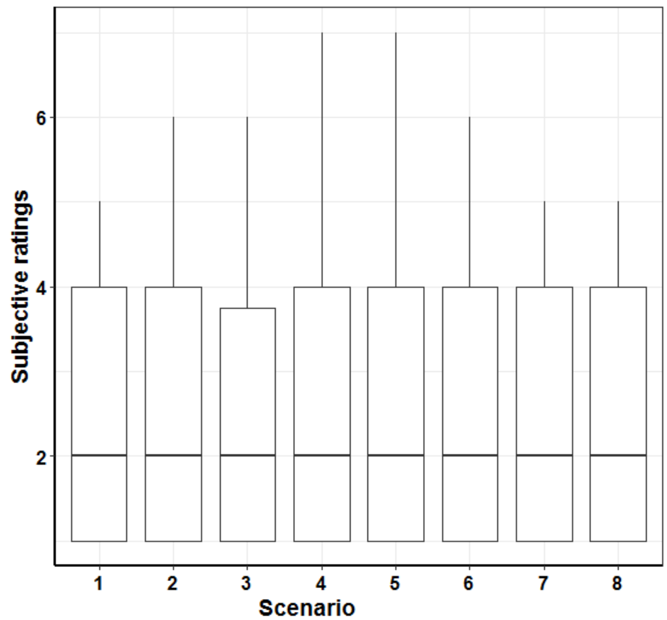


Figure 30. Box plot for subjective rating results

4.4 Statistical analysis

The summary of statistical analysis results of each measurement is illustrated in Table 6. In this table, the first column represents all measurements in the study, including performance score, physiological changes from participants during the experiment (calculated by four different methods: 3-feature LDA, 3-feature SVM, full-feature LDA, and full-feature SVM), and subjective ratings of stress. The second column includes results from repeated measures of ANOVA, which is presented in terms of whether there was a significant difference among scenarios or not. The next column represents the number of significantly different pairs of scenarios from the Tukey post-hoc test. The fourth column shows results for factorial analysis, in the form of which factor or interaction had a significant effect on the responses. The second last column shows correlation analysis results. This column contains the measurement that has significant correlation with the measurement in the first column. Presence of a time effect on the participants' physiological changes is indicated in the final column. The results of these analyses are described in more detail in the following sections.

Table 6. Summary of statistical analysis

Measurement	Significant difference among scenarios	Number of significant different pairs from Tukey test	Factor(s) had significant effect	Significant correlation with	Presence of time effect
Performance score	Yes	7	situation familiarity	Full-feature SVM	
Classification accuracy from:					
• 3-feature LDA	No	None	None	None	No
• 3-feature SVM	Yes	None	Situation familiarity	Subjective ratings of stress	Yes
• full-feature LDA	No	None	Situation familiarity	None	Yes
• full-feature SVM	Yes	None	Situation familiarity	Subjective ratings of stress; performance score	Yes
Subjective ratings of stress	No	None	None	Full-feature SVM	

4.4.1 Repeated measures of ANOVA

The summary of results from repeated measures of ANOVA is presented in Table 7. From Table 7, it can be seen that p-values from repeated measures of ANOVA were less than 0.05 for performance score, and classification accuracy from the 3-feature and full-feature SVM. This means there were significant differences among scenarios in terms of participants' performances and their physiological changes determined by 3-feature and full-feature SVM. Meanwhile, the results also show that there were no significant differences among scenarios in terms of participant's subjective ratings of stress and their physiological changes determined by 3-feature and full-feature LDA.

Table 7. Results summary of repeated measures of ANOVA on the measurements

	Performance score	3-feature LDA	3-feature SVM	full-feature LDA	full-feature SVM	Subjective ratings of stress
p-value	6.04×10^{-7}	0.714	0.0004	0.873	0.0005	1

4.4.2 Tukey test

Results from the Tukey post-hoc test are illustrated in Figures 31-36. In each of these figures, the bars represent results for the Tukey test for each pair of scenarios. It can be seen that with eight scenarios, there are 28 comparison pairs. If a bar does not cross the zero line, the corresponding comparison pair shows a significant difference. In Figures 31-36, there are seven bars that cross the zero line, and they are all in Figure 31, which depicts the results for the performance score. This means only the measurement of the performance score had pair-wise differences among scenarios, which are between scenarios: 8-1, 4-2, 6-2, 8-2, 8-3, 8-5, and 8-7.

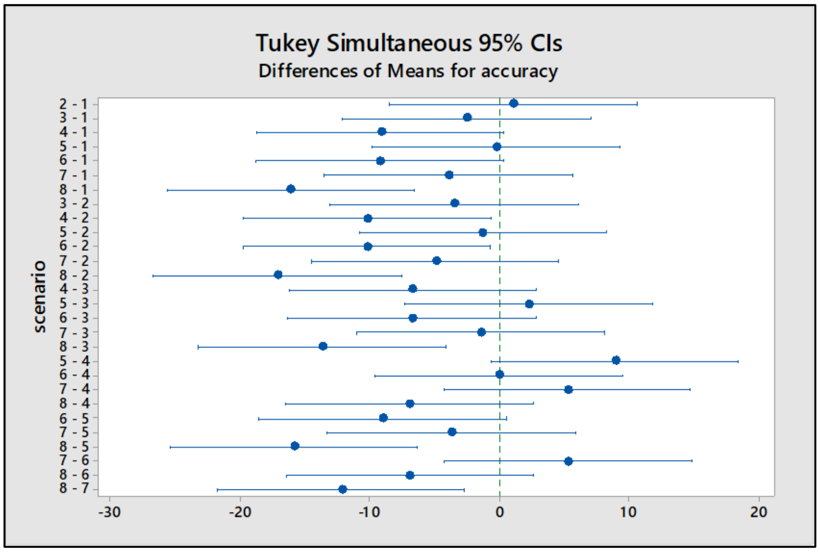


Figure 31. Results from Tukey test for difference among scenarios in terms of performance scores

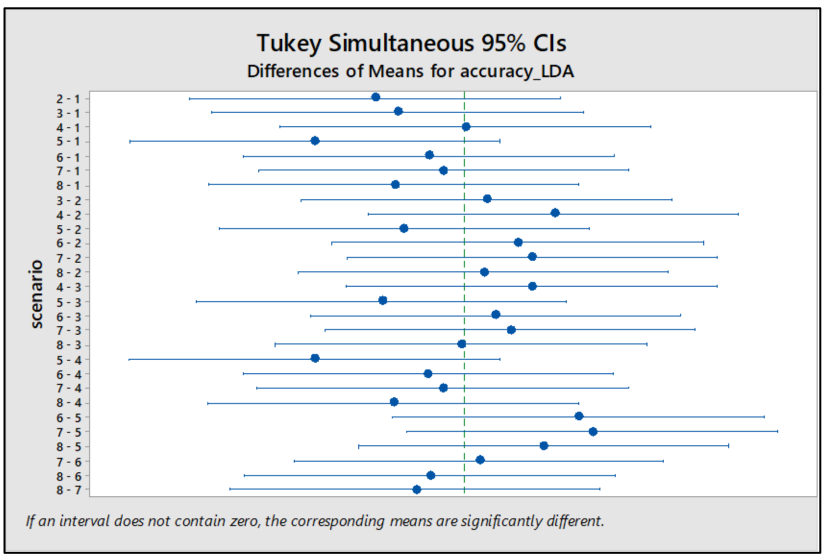


Figure 32. Results from Tukey test for difference among scenarios in terms of 3-feature LDA classification accuracy

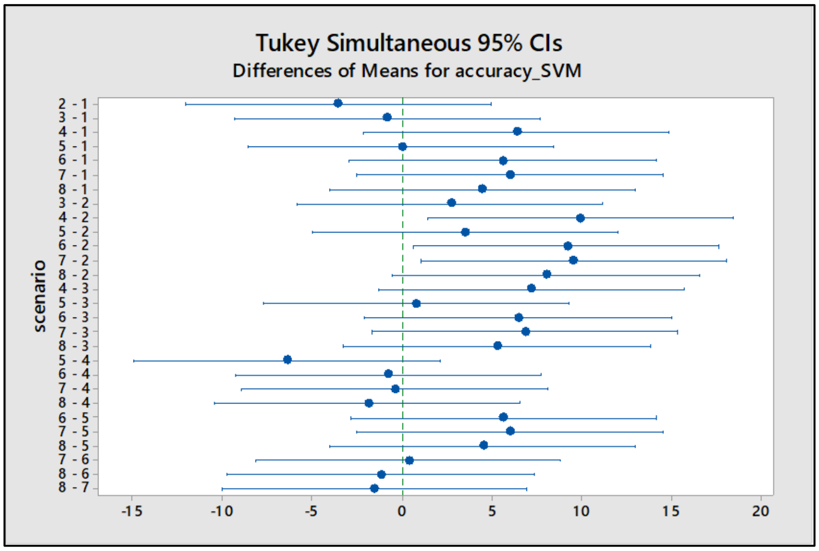


Figure 33. Results from Tukey test for difference among scenarios in terms of 3-feature SVM classification accuracy

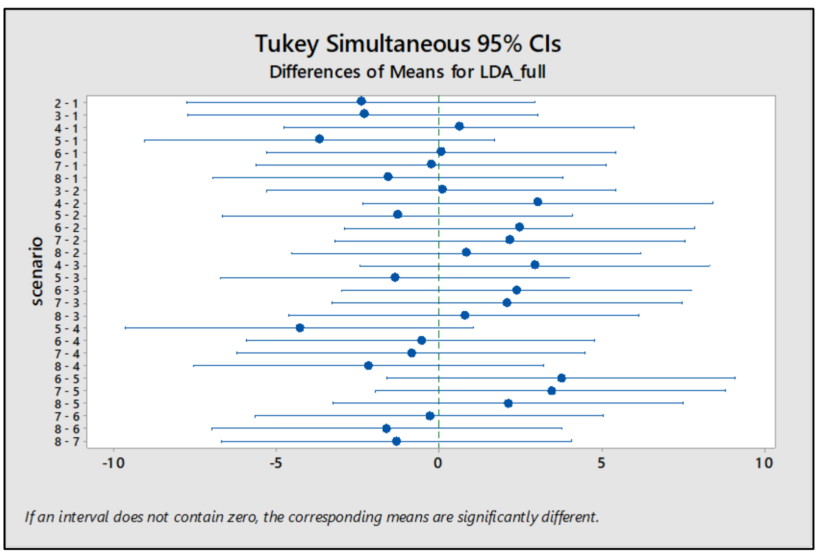


Figure 34. Results from Tukey test for difference among scenarios in terms of full-feature LDA classification accuracy

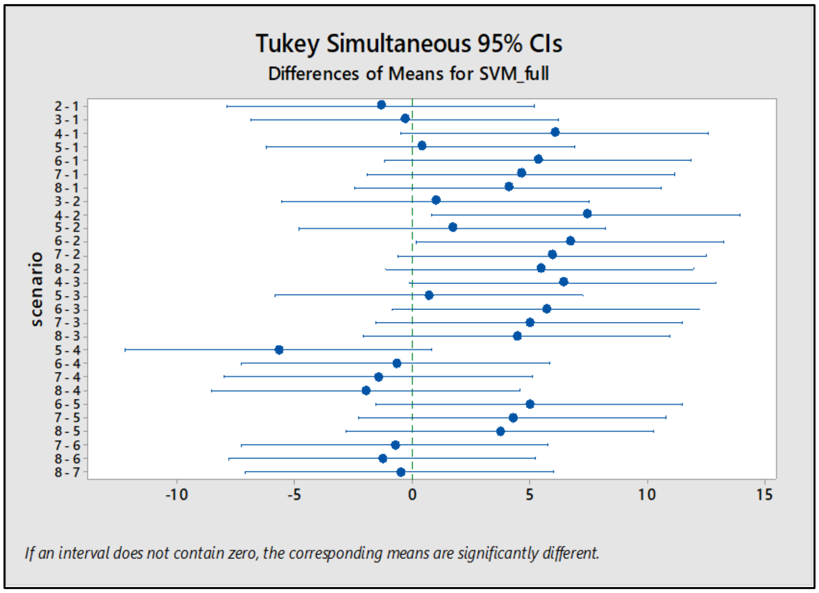


Figure 35. Results from Tukey test for difference among scenarios in terms of full-feature SVM classification accuracy

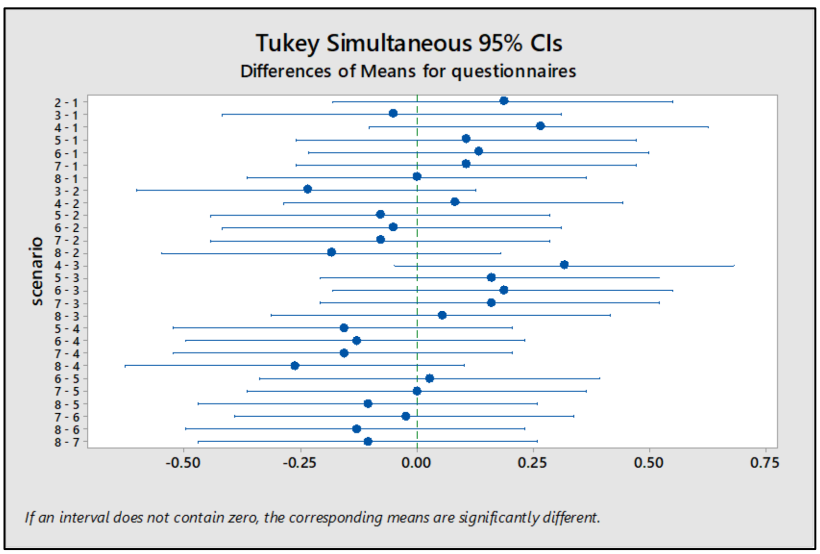


Figure 36. Results from Tukey test for difference among scenarios in terms of subjective ratings of stress

4.4.3 Factorial analysis

The results of factorial analysis are derived from Design Expert 10 and are summarized in Table 8, where p-values for the factors that have significant effect on the response are provided. In this table, it can be noticed that situation familiarity is the only factor that had a significant effect on the responses. Performance score, classification accuracy from full-feature LDA, and 3-feature and full-feature SVM are the responses that were significantly affected by situation familiarity, while classification accuracy from 3-feature LDA and subjective ratings of stress were not affected by any factor.

Table 8. Summary of results from factorial analysis

Response	Factors that had significant effect	p-value
Performance score	Situation familiarity	< 0.0001
Classification accuracy from:		
• 3-feature LDA	None	
• 3-feature SVM	Situation familiarity	< 0.0001
• Full-feature LDA	Situation familiarity	0.027
• Full-feature SVM	Situation familiarity	< 0.0001
Subjective ratings of stress	None	

4.4.4 Correlation analysis

The summary of correlation analysis results is provided in Table 9, where X represents performance score, Y represents subjective ratings of stress, Z_1 represents 3-feature LDA classification accuracy, Z_2 represents 3-feature SVM classification accuracy, Z_3 represents full-feature LDA classification accuracy, and Z_4 represents full-feature SVM classification accuracy.

Table 9. Results summary of correlation analysis

Correlation pair	XY	XZ_1	YZ_1	XZ_2	YZ_2	XZ_3	YZ_3	XZ_4	YZ_4
Correlation rho	-0.272	-0.162	0.246	-0.262	0.405	-0.249	0.253	-0.320	0.353
p-value	0.099	0.330	0.136	0.112	0.012	0.132	0.126	0.050	0.030
Significant	No	No	No	No	Yes	No	No	Yes	Yes

4.4.5 Linear regression

Linear regression was conducted to find any possible trend in the results for classification accuracy from 3-feature LDA, 3-feature SVM, full-feature LDA, and full-feature SVM, in time order. The results are presented in Table 10 and Figure 37. As can be seen from Table 10, only classification accuracy from 3-feature LDA was independent of time; all results from the other three methods were affected by the order of the scenarios.

Table 10. Summary of linear regression results for time-order classification accuracy

	Coefficient	p-value	Inference
3-feature LDA	-0.386	0.136	Insignificant trend
3-feature SVM	-0.946	0.013	Significant decreasing trend
Full-feature LDA	-0.478	0.043	Significant decreasing trend
Full-feature SVM	-0.738	0.014	Significant decreasing trend

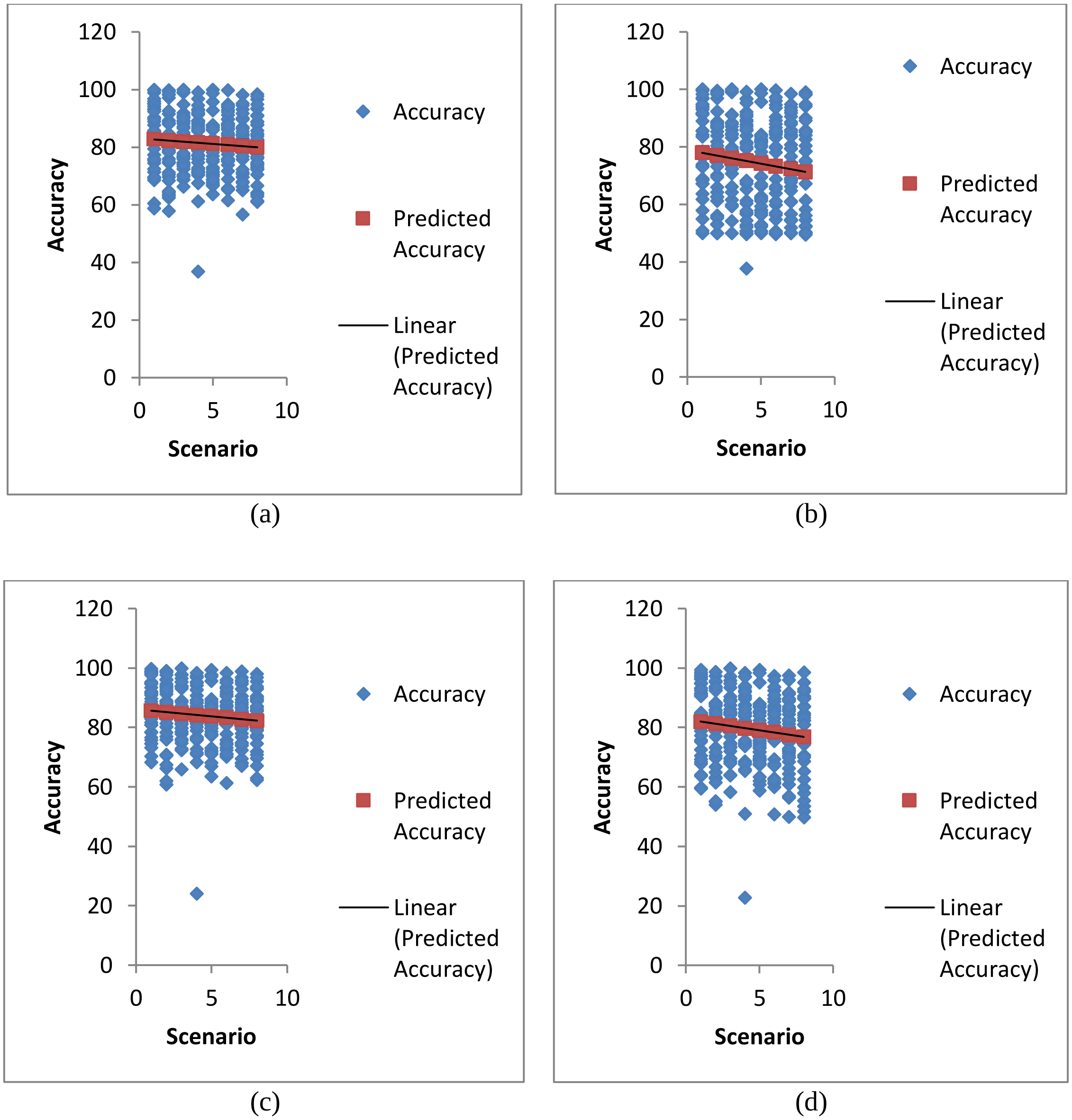


Figure 37. Line fit plot of classification accuracy from (a) 3-feature LDA (b) 3-feature SVM (c) full-feature LDA (d) full-feature SVM

4.5 Modification of physiological classification method

For statistical analysis, it can be noticed that there are some discrepancies among classification accuracies calculated from different methods. Firstly, in repeated measures of ANOVA, while there were significant differences among scenarios in terms of classification accuracy calculated from 3-feature and full-feature SVM, there were no significant differences found among scenarios in terms of classification accuracy calculated from 3-feature and full-feature LDA. Secondly, in factorial analysis results of classification accuracy, only 3-feature LDA showed no significant effect of any factor or interaction on the response, while all the other three methods revealed that situation familiarity had a significant effect on physiological arousals. A similar situation happens in the results of the linear regression, where three methods (3-feaure SVM, full-feature LDA, and full-feature SVM) derived the same results of a significant decreasing trend, the 3-feature LDA's results suggested that there was no trend. Finally, in terms of correlation analysis, physiological measures from LDA methods had no significant correlations with either performance or subjective ratings of stress, whereas, physiological measures from 3-feature SVM had a statistically significant correlation with subjective ratings of stress, and physiological measures from full-feature SVM had a statistically significant correlation with both performance scores and subjective ratings of stress. There was only one test where physiological measures from all methods derived similar results, which was the Tukey-test, where results from all physiological measuring methods show no significantly different pairs of scenarios. From the aforementioned discrepancies, one method should be selected for measuring physiological arousal as the most reliable one to continue discussing. In this case, a joint method was applied (Figure 38). This method was validated by (Smets, et al., 2016; Xu, et al., 2015), indicating that because of the different response of each person to stress, then their data's pattern varies from person to person and a classification method that best fits this person at this time might not best fit him/her at another time, or a classification method that best fits one person might not work well when applied to another person.

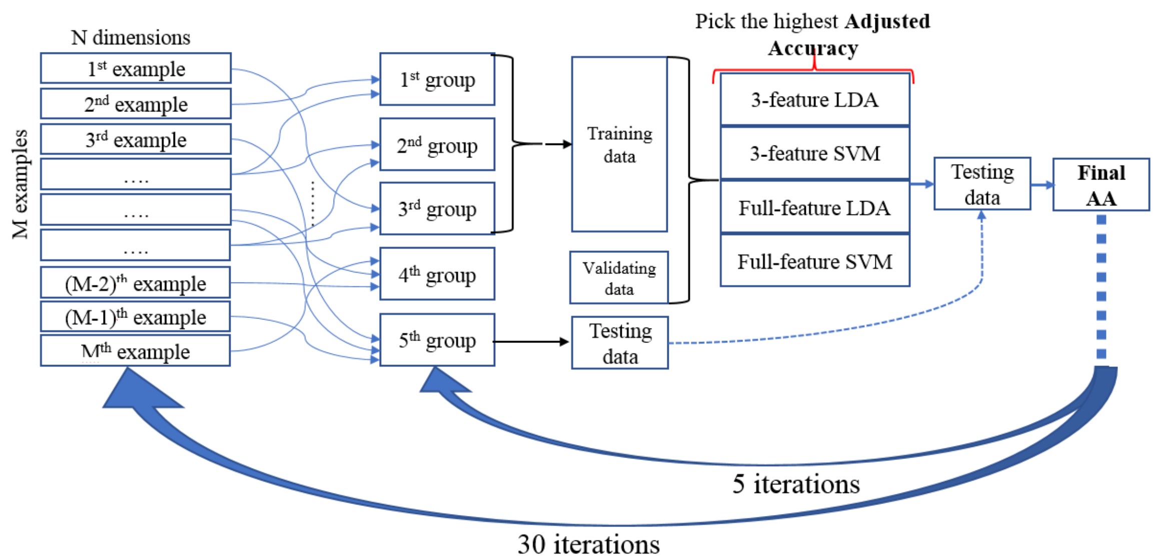


Figure 38. Modification of physiological classification method

Following this method, the algorithms were combined into one where in every group of the cross-validation the adjusted accuracy was calculated by four algorithms and the best one was selected to contribute to the overall classification accuracy. The final results for physiological measures are provided in Figure 39 and Table 11. The average of classification accuracy across participants derived by this algorithm is between 79.6% and 84.5%, with mean 82.4% across scenarios.

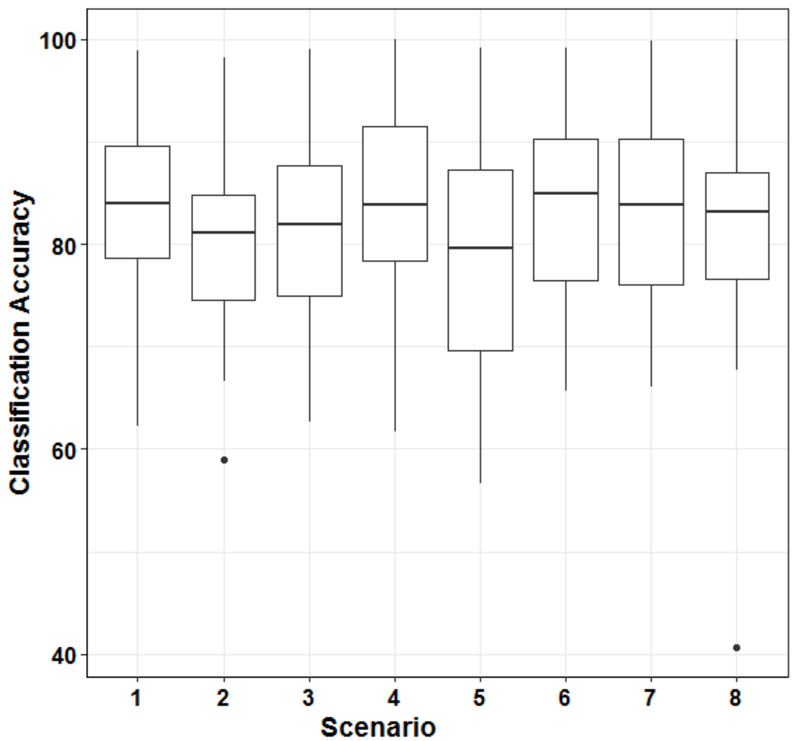


Figure 39. Box plot for classification accuracy from the combined algorithm

Table 11. Classification accuracy from the combined algorithm

Subject	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	Scenario 7	Scenario 8
1	77.93	58.93	75.39	77.27	68.40	74.79	88.98	96.43
2	98.84	93.27	92.83	88.83	99.18	90.14	90.21	92.08
3	79.14	68.40	66.01	61.77	68.87	74.80	73.83	73.89
4	84.02	73.18	86.78	90.95	88.02	88.48	84.34	92.38
5	84.05	74.41	79.45	77.98	80.48	81.48	86.31	82.67
6	78.57	87.53	87.94	87.36	87.28	78.49	92.85	83.88
7	86.47	80.16	82.46	80.23	67.05	83.88	72.49	83.85
8	88.91	93.26	81.45	88.85	96.33	73.25	75.23	82.68
9	89.57	92.67	74.73	89.79	74.41	76.89	66.12	71.01
10	91.39	79.82	76.15	92.62	67.54	85.53	83.99	76.80
11	85.73	84.66	85.64	84.10	83.39	88.70	85.94	74.77
12	82.70	76.07	72.37	87.70	56.72	97.22	83.87	85.23
13	66.54	70.21	91.80	94.58	85.53	77.38	90.42	85.72
14	69.88	83.06	79.84	80.73	82.98	76.36	95.15	75.61
15	96.67	98.22	98.93	99.97	96.13	98.25	99.76	99.58

16	84.50	68.58	69.40	79.47	83.49	72.62	79.28	76.03
17	91.65	81.08	70.62	97.28	75.28	87.47	93.42	76.57
18	89.61	97.47	86.01	85.91	87.17	93.13	88.65	81.06
19	97.34	94.22	94.15	99.98	97.79	99.11	99.62	100.00
20	62.27	75.23	65.11	67.98	62.75	65.70	82.40	82.95
21	78.56	80.34	79.73	93.30	78.48	91.77	90.19	87.43
22	83.92	82.37	62.66	82.45	66.56	72.29	74.54	67.70
23	77.03	81.64	65.62	76.92	86.32	79.02	91.45	82.53
24	73.83	82.39	80.17	83.58	72.04	77.90	74.47	81.98
25	66.38	84.83	74.24	77.12	62.72	84.98	71.00	69.03
26	83.25	72.63	83.48	71.65	78.62	72.47	81.56	83.23
27	89.53	66.61	65.12	74.76	69.75	89.30	71.12	99.47
28	78.81	75.03	80.30	98.58	97.23	72.16	78.62	40.73
29	86.90	83.31	89.03	83.30	69.89	80.46	83.27	83.83
30	81.37	70.63	89.94	79.38	81.64	87.42	83.86	90.10
31	80.10	81.07	86.82	75.93	73.30	72.14	68.85	83.97
32	92.22	80.88	85.77	94.66	95.62	84.77	72.55	73.37
33	83.43	84.57	86.05	91.62	69.82	91.79	83.34	81.65
34	82.20	94.82	83.85	93.94	98.51	94.65	91.33	98.05
35	77.03	76.59	95.55	72.89	65.58	85.98	94.52	84.40
36	93.72	91.37	90.52	82.57	85.10	92.02	82.81	91.93
37	85.21	70.54	77.16	80.13	69.56	90.29	80.70	81.78
38	90.31	84.64	92.52	84.00	94.22	92.70	85.95	84.40
Average	**83.41**	**80.91**	**81.20**	**84.48**	**79.57**	**83.57**	**83.50**	**82.60**

Updated statistical analysis results for the measures of physiological arousals is presented in Table 12.

Table 12. Summary of results from statistical analysis of physiological changes calculated by the combined algorithm

Measurement	Significant difference among scenarios	Number of significant different pairs from Tukey test	Factor(s) had significant effect	Significant correlation with	Availability of time effect
Physiological changes calculated from the combined algorithm	No	None	Situation familiarity	Subjective ratings of stress (marginally)	Yes

From repeated measures of ANOVA, the p-value was found to be 0.564, suggesting that there was no significant difference among scenarios in terms of participants' physiological changes. Results from Tukey test are provided in Figure 40, indicating that there are no pairs of scenarios having a significant difference. Meanwhile, factorial analysis results suggested that situation familiarity is the only factor that had a significant effect on physiological measure (p-value = 0.0176). In addition, correlation analysis results show that physiological changes had no significant correlations with either performance scores ($\rho = -0.251$, $p = 0.128$) or subjective ratings of stress ($\rho = 0.303$, $p = 0.064$). Linear regression results indicate that participants' physiological changes were affected by time factor, with the classification accuracy decreasing over the course of the experiment (p-value = 0.045, see Figure 41 for regression plot).

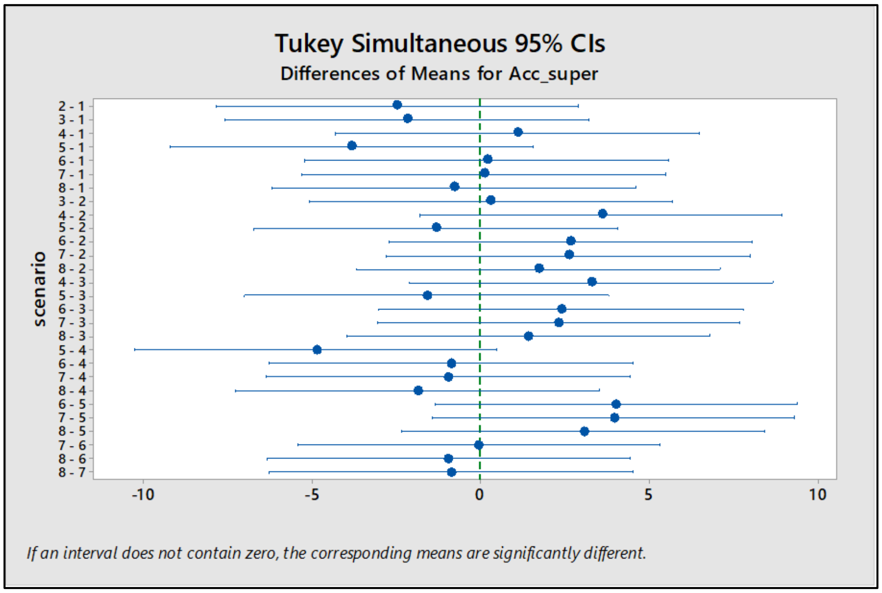


Figure 40. Results from Tukey post-hoc test for comparing participant's physiological changes among scenarios

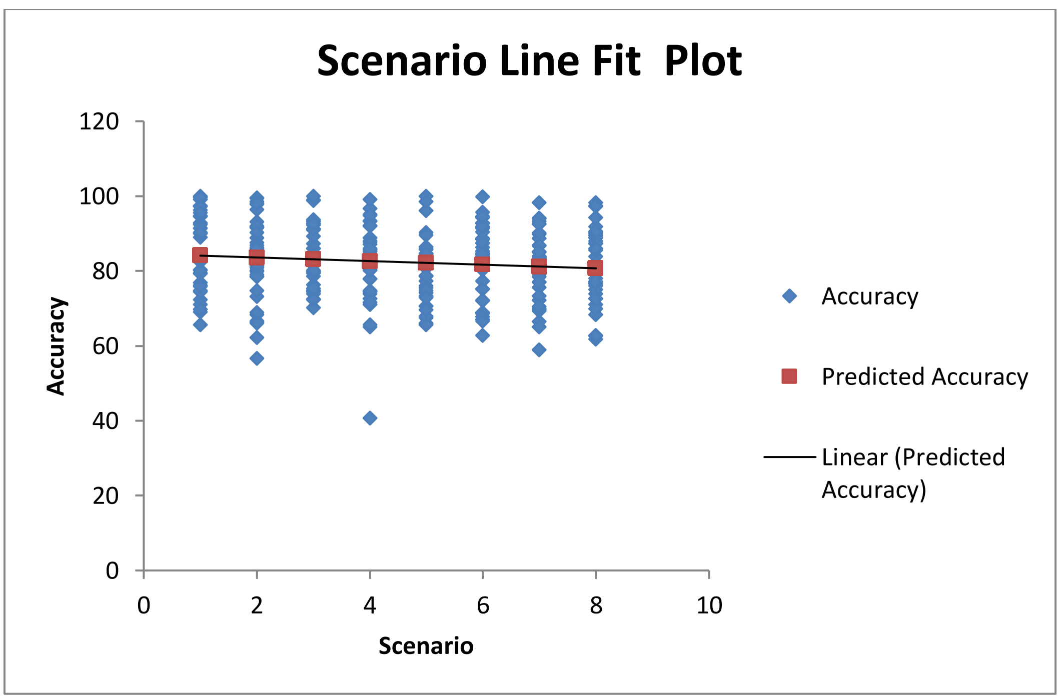
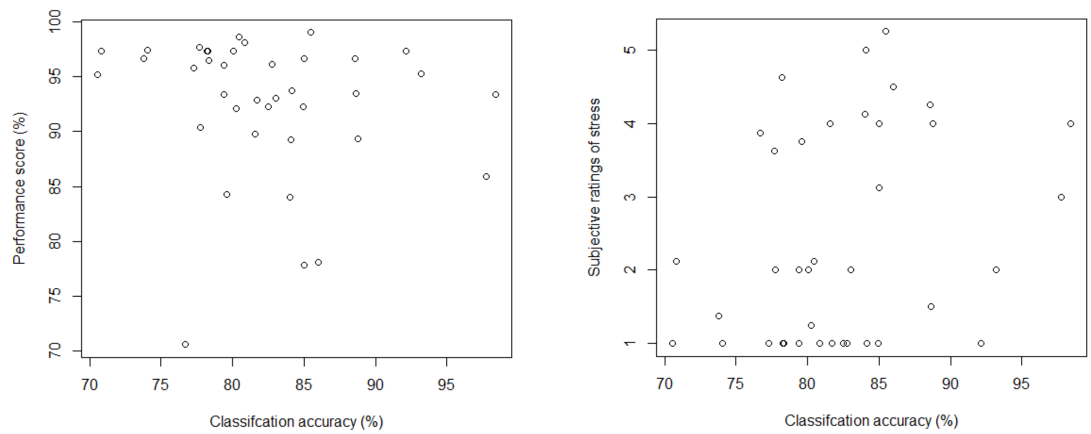


Figure 41. Regression plot of physiological changes in participants in time order

Chapter 5: Discussion

In all scenarios, the classification accuracy of the ANS signals were significantly different from chance (mean 81.3%), indicating that all scenarios were effective in eliciting a physiological response. Furthermore, the classification accuracies were not correlated with performance scores ($\rho = -0.251$, $p = 0.128$), supporting the idea that this measure could provide results that performance scores alone do not reveal. This supports the hypothesis that was posed in the introduction: in training, especially in emergency situations, some people might derive the same results (performance scores), but their mental states during the situation might be different. From Figure 42(a) (scatter plot), most participants derived high performance in terms of scoring (mostly over 80%), but their physiological arousal levels vary over a wide range (from 68% to 99% of classification accuracy). Note that there are 38 data points in the figure. Those data points are the average of 8 data points from 8 scenarios for each participant. There are participants who had similar levels of physiological changes during training, but their performance was noticeably different. This is reasonable because it has been proved that high physiological arousals could mean either higher engagement in the environment or higher stress (Patton & Gamble, 2016; Lackey, et al., 2016; Bian, et al., 2015). While stress is related to a decrease in performance (Winslow, et al., 2015), environment engagement leads to better concentration, thus increasing the participant's outcome (Patton & Gamble, 2016).

Physiological arousal levels were marginally correlated to the results from subjective ratings of stress ($\rho = 0.303$, $p = 0.064$). As can be seen from Figure 42(b), there is a pattern in the data, although not quite clear, suggesting there is a positive relation between subjective ratings of stress and physiological arousals measured objectively. This supports the claim that physiological measures can detect participants' stress.



(a) (b)

Figure 42. Scatter plot of classification accuracy from physiological classification accuracy and (a) performance score (b) subjective ratings of stress

Results from the factor analysis indicate that familiarity with the environment was the only factor that had a significant effect on the difficulty of the task and the user experience. This factor showed a significant effect on both the trainees' performance scores ($p < 0.0001$$p < 0.0001$) and their level of physiological arousal ($p = 0.0176$ $p < 0.0001$), but not on the subjective ratings of stress. There were no significant interaction effects seen in any of the measures. Also, repeated measures of ANOVA showed significant differences in only the performance score ($p = 6.04 \times 10^{-7}$), further indicating that the scenarios are different in difficulty levels (measured by performance scores), but not in participants' experiences (measured by physiological arousal). This outcome re-assures us that physiological measures provide information about a scenario that may not be seen in performance scores.

Although there was no significant difference in participants' experiences on average (proved by repeated ANOVA results), there were differences in participants' experiences individually. In other words, one participant might find a scenario more stressful than another, while another participant might find the opposite. For example, from Table 11, it can be seen that subject 1 had the physiological arousal level of approximately 59% for scenario 2, the lowest among other scenarios. Whereas, subject 9 found scenario 2 the most stressful compared to other scenarios, with 92.8% physiological arousal level. Or in another case, while participant 2 found scenario 6 the most stressful with 97.2% of physiological changes, participant 28 found it less stressful with only 72.2% of stress scale. Taking a closer look, it can be seen that participants' physiological responses to a training scenario vary from one participant to another. Figure 43 depicts the histogram of the number of scenarios whose physiological changes are in certain ranges in each participant. It can be noticed that some participants have high stress levels in many of scenarios, while some other participants have the opposite pattern. This confirms that participants' experiences were different from each other, although they did the same training scenario.

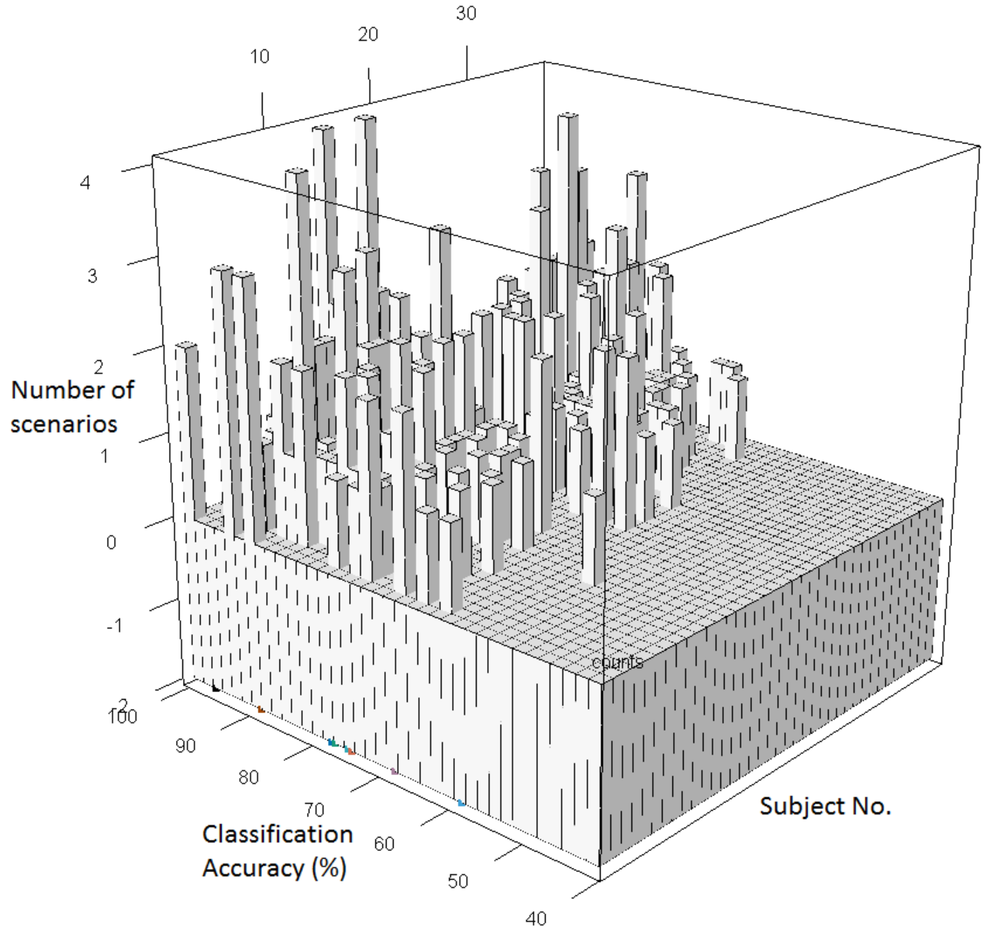


Figure 43. Histogram of number of scenarios whose physiological changes in certain ranges in each participant

At the beginning of the experiment design, we predicted that all three performance shaping factors would have a significant effect on the responses. However, the results showed that only familiarity with the environment significantly affected the participants' performance and their level of physiological arousal or stress. There might be several reasons for this. For example, the gap between high and low levels of PSF 1 (quality of information received during scenario) and PSF 2 (proximity to hazard) may not have been significant enough to make a difference, or those two factors might merely not be significant and should be ignored when designing training scenarios in the future.

In the results from case 1 of k-means clustering, it could be noticed that baseline data disperse quite equally over four clusters, while scenarios' data fluctuate from cluster to cluster (Figure 22). Another point to be noticed is that the boxes for baseline data are much smaller than the ones for scenario data, suggesting that baseline data do not vary as much as scenario data do over different participants (Figure 22). These two points support the idea that participants' mental states are more stable during baseline sessions than during training sessions. Similarly, results from case 1 of Gaussian model clustering also show that the proportion of anomaly baseline data points varies slightly across different participants, while this amount varies much more in the case of scenario data (illustrated by a small box for baseline data and big boxes for scenario data in Figure 26). Other than this, there is no clear pattern seen from clustering results.

Results from linear regression showed that there is a trend of physiological changes over the order of scenarios that participants did. This is a downward trend, indicating that participants' mental states vary more in the first scenarios and became more stable when the session proceeded to the end. This could be explained by the fact that participants were calmer when they were familiar with the platform. This point could not be seen from performance score, which once again indicates the capability of physiological measures to reveal information of trainees' experiences during a training session.

For the aforementioned points, an objective measure of stress like physiological arousal may be helpful in estimating participants' feelings while in a VE- or simulation-based training. For example, it could be used as an additional indicator of competency in the trained skills to complement the performance measure. In addition, such a measure could be a tool to evaluate a person's capability to work offshore. For example, different participants could derive the same results in emergency scenarios in terms of performance; their stress levels, however, could be very different. As human failures are highly correlated to stress (Cohen, 1980; Hockey, 1997), trainees who are more prone to an increase in stress have the potential to perform worse in real conditions, where the emergency is real. Therefore, applying stress detection in training might be a solution for organizations who need to choose people with solid performance during critical situations. Finally, the physiological arousal measurement could also be incorporated into VR applications to monitor users' feelings in real-time, allowing for modification of the scenarios accordingly to enhance each individual's experience while using VR applications.

Chapter 6: Conclusions and Future Work

6.1 Conclusions

After applying physiological measurements to participants in 8 different VR-based emergency response training scenarios, it was found that although their performances were different among scenarios, their mental experiences were not. This information could be useful for those who design virtual scenarios for emergency training, especially if they want to create scenarios where trainees will experience different levels of stress (different levels of sense of emergency). Specifically, this study's findings suggest that the scenario designers might need to increase the gaps between the high and low levels of the performance shaping factors, or use other factors in the design, in order to create different levels of sense of emergency for the training program. This would help trainees experience the training program with the difficulty levels from low to high, thus helping them to learn the emergency procedure from basic levels to advanced levels.

In conclusion, classification accuracy between physiological data collected during a training scenario and that collected during baseline can be a useful measure of trainees' experience in a given training scenario to complement performance measures, which is potentially useful for training program designers in designing the curriculum for VR-based training programs. Furthermore, physiological signals may be more reliable indicators of stress than subjective ratings. The findings from this research might be useful in a number of applications.

6.2 Future Work

It is necessary to conduct further research to distinguish between stress and engagement in participants' physiological changes during virtual training, so that a clear relationship between performance and stress, and between performance and engagement could be found. In addition, during training sessions, the questionnaires should include both stress and engagement questions, thus providing more comprehensive responses from participants.

Finally, experiments should be designed to be more different in terms of difficulty levels. From this study's findings, situation familiarity is a good factor to make scenarios different. The other two factors, which are *proximity to hazard* and *information quality* could either be designed to have bigger gaps between low level and high level, or replaced by other potential factors.

References

Ahlberg, G. et al., 2007. Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. *The American journal of surgery,* 193(6), pp. 797-804.

Andreatta, P. et al., 2010. Virtual reality triage training provides a viable solution for disaster-preparedness. *Academic emergency medicine,* 17(8), pp. 870-876.

Aoki, H., Oman, C. & Natapoff, A., 2007. Virtual-reality-based 3D navigation training for emergency egress from spacecraft. *Aviation, space, and environmental medicine,* 78(8), pp. 774-783.

Basdogan, C., Ho, C. & Srinivasan, M., 2001. Virtual environments for medical training: Graphical and haptic simulation of laparoscopic common bile duct exploration. *IEEE/Asme Transactions On Mechatronics,* 6(3), pp. 269-285.

Bian, D. et al., 2015. *Physiology-based Affect Recognition During Driving in Virtual.* Angers, France, Science and Technology Publications, Lda., pp. 137-145.

Block, J. & Burns, R., 1976. Mastery learning. *Review of research in education,* 4(1), pp. 3-49.

Brouwer, A. et al., 2013. Perceiving blocks of emotional pictures and sounds: effects on physiological variables. *Frontiers in Human Neuroscience,* pp. Vol. 7, Article 295.

Carlson, N., 2013. *Physiology of Behavior.* New Jersey, Pearson Education, Inc..

Chang, S., Kwon, E., Kwon, Y. & Kwon, H., 2010. The effects of simulation training for new graduate critical care nurses on knowledge, self-efficacy, and performance ability of emergency situations at intensive care unit. *Korean Journal of Adult Nursing,* 22(4), pp. 375-383.

Chan, J., Leung, H., Tang, J. & Komura, 2011. A virtual reality dance training system using motion capture technology. *IEEE Transactions on Learning Technologies*, 4(2), pp. 187-195.

Cho, D. et al., 2017. Detection of Stress Levels from Biosignals Measured in Virtual Reality Environments Using a Kernel-Based Extreme Learning Machine. *Sensors,* 17(10), p. 2435.

Christopher, M., 2016. *Pattern Recognition and Machine Learning.* New York: Springer-Verlag.

Cohen, S., 1980. Aftereffects of stress on human performance and social behavior: a review of research and theory. *Psychological bulletin,* 88(1), p. 82.

Conesa, J., 1995. Electrodermal palmar asymmetry and nostril dominance. *Perceptual and Motor Skills,* pp. 211-216.

Cristianini, N. & Shawe-Taylor, J., 2000. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods.* 1 ed. Cambridge: Cambridge University Press.

Egan, D. et al., 2016. *An evaluation of Heart Rate and ElectroDermal Activity as an objective QoE evaluation method for immersive virtual reality environments.* Lisbon, Portugal, IEEE.

Erren-Wolters, C. et al., 2007. Virtual reality for mobility devices: training applications and clinical results: a review. *International journal of rehabilitation research,* 30(2), pp. 91-96.

Figner, B. & Murphy, R. O., 2010. Using skin conductance in judgment and decision making research. *A Handbook of Process Tracing Methods for Decision Research: A Critical Review and User's Guide,* pp. 163-84.

Fisher, R. A., 1936. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics,* Volume 7, pp. 179-188.

Gallagher, A. et al., 1999. Virtual reality training in laparoscopic surgery: a preliminary assessment of minimally invasive surgical trainer virtual reality (MIST VR). *Endoscopy,* 31(04), pp. 310-313.

Garcia-Ceja, E., Osmani, V. & Mayora, O., 2016. Automatic Stress Detection in Working Environ-ments from Smartphones' Accelerometer Data: A first step. *IEEE Journal of Biomedical and Health Informatics,* 20(4).

Gu, Q., Li, Z. & Han, J., 2011. *Generalized Fisher score for feature selection.* Barcelona, Spain, AUAI Press Arlington, pp. 266-273.

Gurusamy, K., Aggarwal, R., Palanivelu, L. & Davidson, B., 2009. Virtual reality training for surgical trainees in laparoscopic surgery. *Cochrane Database Syst Rev,* 1(4).

Hockey, G., 1997. Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological psychology,* 45(1-3), pp. 73-93.

House, A., Smith, J., MacKinnon, S. & Veitch, B., 2014. *Interactive simulation for training offshore workers.* St.John's, Oceans, p. DOI: 10.1109/OCEANS.2014.7003211.

Hsu, E. et al., 2013. State of virtual reality based disaster preparedness and response training. *PLoS Currents,* p. DOI: 10.1371/currents.dis.1ea2b2e71237d5337fa53982a38b2aff.

Huysmans, D. et al., 2018. Unsupervised Learning for Mental Stress Detection-Exploration of Self-Organizing Maps. *Scitepress.*

Jones, R., Kazdin, A. & Haney, J., 1981. Social validation and training of emergency fire safety skills for potential injury prevention and life saving. *Journal of Applied Behavior Analysis,* 14(3), pp. 249-260.

Kecman, V., 2001. Learning and Soft Computing. *MIT Press.*

Lackey, S. J., Salcedo, J. N., Szalma, J. & Hancock, P., 2016. The stress and workload of virtual reality training: the effects of presence, immersion and flow. *Ergonomics,* 59(8), pp. 1060-1072.

Loggia, M. L., Juneau, M. N. & Bushnell, M. C., 2011. Autonomic responses to heat pain: Heart rate, skin conductance, and their relation to verbal ratings and stimulus intensity. *Pain,* p. 592–598.

Malik, M. et al., 1996. Heart Rate Variability. *European Heart Journal,* 17(3), pp. pages 354-381.

Mantovani, F., Castelnuovo, G., Gaggioli, A. & Riva, G., 2003. Virtual reality training for health-care professionals. *CyberPsychology & Behavior,* 6(4), pp. 389-395.

Mantovani, F., Castelnuovo, G., Gaggioli, A. & Riva, G., 2003. Virtual Reality Training for Health-Care Professionals. *CyberPsychology & Behavior,* 6(4), pp. 389-395.

McLachlan, G. & Peel, D., 2004. *Finite mixture models.* s.l.:John Wiley & Sons.

Merién, A. et al., 2010. Multidisciplinary team training in a simulation setting for acute obstetric emergencies: a systematic review. *Obstetrics & Gynecology,* 115(5), pp. 1021-1031.

Montgomery, D., 2013. *Design and analysis of experiments.* 8 ed. s.l.:John Wiley & Sons.

Moyle, A. & Veitch, B., 2017. *Assessing Human Participants' Response to an Emergency Situation Using a Virtual Environment as a Diagnostic Tool,* St.John's: Memorial University of Newfoundland.

Nagai, Y., Goldstein, L. H., Fenwick, P. B. C. & Trimble, M. R., 2004. Clinical efficacy of galvanic skin response biofeedback training in reducing seizures in adult epilepsy: A preliminary randomized controlled study. *Epilepsy & Behavior*, pp. 216-223.

Ollander, S., Godin, C., Campagne, A. & Charbonnier, S., 2016. *A comparison of wearable and stationary sensors for stress detection.* Budapest, Hungary, IEEE International Conference on Systems, Man, and Cybernetics (SMC).

Pan, J. & Tompkins, W. J., 1985. A Real-Time QRS Detection Algorithm. *IEEE Transactions on Biomedical Engineering,* BME-32(3).

Panju, S. et al., 2015. Atypical sympathetic arousal in children with autism spectrum disorder and its association with anxiety symptomatology. *Molecular Autism,* 6(64).

Patton, D. & Gamble, K., 2016. Physiological Measures of Arousal During Soldier-Relevant Tasks Performed in a Simulated Environment. *Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience,* Volume 9743, pp. 372-382.

Picard, R. W., Vyzas, E. & Healey, J., 2001. Toward Machine Emotional Intelligence: Analysis of Affective Physiological State. *IEEE Transactions on pattern analysis and machine learning,* 23(10).

Plarre, K. et al., 2011. *Continuous inference of psychological stress from sensory measurements collected in the natural environment.* Chicago, IL, USA, IEEE 10th International Conference on Information Processing in Sensor Networks (IPSN).

Rizzo, A. et al., 2009. A virtual reality scenario for all seasons: the virtual classroom. *Cns Spectrums,* 11(1), pp. 35-44.

Rizzo, A. et al., 2012. Stress resilience in virtual environments: training combat relevant emotional coping skills using virtual reality. *Proceedings of the 9th Conference on Disability, Virtual Reality & Associated Technologies.*

Rosenbaum, M., Creedon, D. & Drabman, R., 1981. Training preschool children to identify emergency situations and make emergency phone calls. *Behavior Therapy,* 12(3), pp. 425-435.

Roth, D., Latoschik, M., Vogeley, K. & Bente, G., 2015. Hybrid Avatar-Agent Technology–A Conceptual Step Towards Mediated "Social" Virtual Reality and its Respective Challenges. *i-com,* 14(2), pp. 107-114.

Scholkopf, B. & Smola, A., 2002. Learning with Kernels. *MIT Press.*

Seymour, N. et al., 2002. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Annals of surgery,* 236(4), p. 458.

Sharma, N. & Gedeon, T., 2012. Objective measures, sensors and computational techniques for stress recognition and classification: A survey. *Computer methods and programs in biomedicine,* 108(3), pp. 1287-1301.

Shaw, L. et al., 2015. Challenges in virtual reality exergame design.

Smets, E. et al., 2016. Comparison of Machine Learning Techniques for Psychophysiological Stress Detection. *MindCare,* Volume DOI: 10.1007/978-3-319-32270-4_2, pp. 13-22.

Smith, J. & Veitch, B., 2015. *The Effect of Virtual Environment Training on Participant Competence and Learning in Offshore Emergency Egress Scenarios,* St.John's: Memorial University of Newfoundland.

Suykens, J. et al., 2002. Least Squares Support Vector Machines. *World Scientific.*

Tovian, S. et al., 2018. *American Psychological Association.* [Online]
Available at: http://www.apa.org/helpcenter/stress-body.aspx
[Accessed 11 9 2018].

Van Wyk, E. & De Villiers, R., 2009. Virtual reality training applications for the mining industry. *Proceedings of the 6th international conference on computer graphics, virtual reality, visualisation and interaction in Africa,* pp. 53-63.

Vrijkotte, T., Doornen, V., L.J. & De Geus, E., 2000. Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability. *Hypertension*, 35(4), pp. 880-886.

Waycott, J. et al., 2018. Manipulating Reality?: Designing and Deploying Virtual Reality in Sensitive Settings. *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 411-414.

Winslow, B. D. et al., 2015. Identification of resilient individuals and those at risk for performance deficits under stress. *Frontiers in Neuroscience.*

Xu, Q., Nwe, T. L. & Guan, C., 2015. Cluster-Based Analysis for Personalized Stress Evaluation Using Physiological Signals. *IEEE Journal of Biomedical and Health Informatics,* 19(1), pp. 275-281.

Yerkes, R. M. & Dodson, J. D., 1908. The relation of strength of stimulus to rapidity of habit-formation. *ournal of Comparative Neurology and Psychology,* pp. 459-482.

Yoder, C. N., 2011. *Mathworks.* [Online]
Available at: http://www. mathworks. com-/matlabcentral/fileexchange/25500
[Accessed 7 6 2018].

Zeng, F., Yap, R. & Wong, L., 2002. Using feature generation and feature selection for accurate prediction of translation initiation sites. *Genome Informatics,* Volume 13, pp. 192-200.

Appendices

Appendix A: Recruiting information

Consent Form (Moyle & Veitch, 2017)

Informed Consent Form

Title: Assessing Human Participants' Response to an Emergency Situation Using a Virtual Environment as a Diagnostic Tool

Researcher(s):

<table>
<tr><td colspan="2">Principle Investigators</td></tr>
<tr><td>Dr. Brian Veitch
Engineering and Applied Science
MUN
(709) 864-8970
bveitch@mun.ca</td><td>Ms. Allison Moyle
Engineering and Applied Science
MUN
(709) 685-5793
p13dabm@mun.ca</td></tr>
<tr><td colspan="2">Co-Investigators</td></tr>
<tr><td colspan="2">Ms. Jennifer Smith | Mr. Kyle Doody |
Engineering and Applied Science
(709) 864-6764
jennifersmith@mun.ca | kdoody@mun.ca |

Dr. Sarah Power
Faculty of Engineering and Applied Science
Faculty of Medicine
(709) 864-8200
b09sdp@mun.ca</td></tr>
</table>

You are invited to take part in a research project entitled *"Assessing human participants' response to an emergency situation using a Virtual Environment as a diagnostic tool"*

This form is part of the process of informed consent. It should give you the basic idea of what the research is about and what your participation will involve. It also describes your right to withdraw from the study and its risks and benefits to be able to make an informed decision. . Take time to read this carefully and to understand the information given to you. Please contact the researcher, *Allison Moyle* , or co-investigator *Jennifer Smith* if you have any questions about the study or would like more information before you consent. Contact information is listed above.

Introduction:
We are an interdisciplinary research team consisting of faculty, staff, and students at Memorial University. This research project is funded jointly by NSERC, Husky Energy and RDC.

Purpose of study:
This study will involve a Human Reliability Assessment (HRA) to assess human performance results in emergency scenarios using a virtual environment. As you have learned in the initial research study and the retention study, AVERT is a virtual environment program that enables users to learn basic offshore emergency response skills through a series of learning objectives, training materials, and evaluations. Training using AVERT is beneficial as it gives individuals access to realistic training scenarios that they wouldn't be exposed to otherwise due to ethical, financial, and logistical constraints.

The objective of this research is to determine if AVERT is an effective way to evaluate how people will react in an offshore safety emergency situation. This will determine the usefulness of AVERT as a diagnostic tool to evaluate participant's strengths and weaknesses in different evacuation conditions to suggest further training in order for an individual to become competent effectively.

What you will do in this study:
You will attend one session at the Virtual Environments (VE) Lab. You will be given an explanation of the experimental design, given an opportunity to ask questions or express concerns, and, if satisfied, will indicate your free and informed consent by completion of this Informed Consent form.

Testing:
Performance Shaping Factors (PSFs) contribute to an event in a positive or negative way and can increase or decrease human performance. PSFs have been incorporated into

various scenarios in high and low levels. Scenarios will involve completing various tasks during an emergency situation using the knowledge you have gained during the AVERT training. Participants will be asked to complete eight scenarios that will have different levels of PSFs.

Collecting Physiology:

Before starting the scenarios, you will have sensors applied to locations on the head, torso and hand. A five-minute seated baseline of physiological signals will be collected prior to the start of each scenario. You will be asked to refrain from exercise, smoking and caffeine for four hours prior to testing, to refrain from alcohol for 24 hours prior to testing, and not to have fasted for a period greater than 2 hours. You will also be asked to wear comfortable clothes.

Length of time:

You will be asked to attend one session. The total time to complete the session is expected to be 1-2 hours (depending on individual performance).

Withdrawal from the study:

If you decide to withdraw from the study, the information collected up to that time will be removed from the study. This information will be destroyed and will not be included in the data analysis of the study.

If you choose to withdraw from the study after data collection has ended, your data can be removed from the study up to two weeks after the completion of your participation.

Possible benefits:

There are no known direct benefits to the participants of this study. With regards to the community, the outcomes of this research may support efforts that improve training of maritime and offshore industry personnel, and thereby contribute to an improvement in the safety of those industries.

The findings from this study will advance knowledge. Specifically, the findings will determine if a virtual environment can be used as a diagnostic tool for assessing performance during an offshore emergency situation. The findings may also inform improvements to safety training that result in safer industry practices, which is a societal benefit.

Possible risks:

If you are not comfortable with any aspect of the testing, then you have the right to withdraw from the study at any point.

Navigation through the virtual space using a desktop computer configuration may cause some individuals to experience symptoms of visually induced motion sickness (VIMS) or simulator-induced sickness (SIS). The symptoms of simulator-induced sickness include fatigue, headache, eye strain, difficulty focusing, increased salivation, sweating, nausea, stomach awareness, blurred vision, dizziness, vertigo and burping. The symptoms of simulator sickness can sometimes occur during, immediately after or several hours after exposure to the simulator.

To ensure you do not experience severe symptoms, simulator-induced sickness susceptibility will be assessed prior to the study and will be monitored throughout using the simulator sickness questionnaire (SSQ). The research coordinator will monitor you during the trials for symptoms and stop the trials if necessary. The simulator sickness questionnaire allows you to rate the severity of your symptoms as no symptoms, minimal, moderate and severe. A SSQ will be completed after the first four scenarios and again at the end of the scenario testing. If you self-report a symptom as moderate, then the trials will be paused and you will be provided an extended rest period to allow symptoms to subside until you are able to proceed. If you report a symptom as severe, the trials will be stopped and you will be provided with a rest period until symptoms subside. Should any symptoms persist (beyond a period of 20 minutes), you will be excluded from the study.

To reduce the effects of simulator-induced sickness, your exposure time to the virtual environment will be limited to a maximum of 10 minutes per scenario with time allocated for breaks in between scenarios to allow a period of rest. It may be unsafe to drive if symptoms persist after the rest period. If symptoms persist, arrangements will be made to take you home. Symptoms must subside before you are able to leave the experimental laboratory.

Exposure to virtual reality may cause seizures for some individuals. Individuals who are prone to seizures or have a history of seizures will not be eligible to participate in the study. You will be monitored throughout the study to ensure you do not experience seizures. The research team will be trained in standard First Aid should a situation arise.

Exposure to a desktop computer screen may cause eye strain in some participants. Screen time exposure is minimal, and therefore there is minimal expected discomfort. The distance from you to the screen will be selected such that it reduces the potential for eye strain and discomfort. Eye strain is expected to be not more than would be experienced during normal computer usage of the same duration.

Electrodes/sensors will be applied at the following locations: hand and torso (rib caged area). While these are only applied to the skin, the conductive gel that is used to ensure signal quality, and tape that is used to secure the wires, may irritate sensitive skin. The application method employed in this study is common practice in research and clinical applications. Skin sensitivity will be assessed prior to the application of the sensors and should the skin become irritated to a point of discomfort you retain the right to withdraw. All efforts will be made to minimize the duration of skin exposure to the adhesive gel and tape.

Performance in the virtual environment scenarios will be assessed repeatedly throughout the study. For some individuals, this may cause performance anxiety or stress which may cause poor performance in the test scenarios. To reduce the likelihood of anxiety and stress, you will receive a break between stages to rest and be instructed not to worry or dwell on the previous testing scenarios.

Some participants may experience embarrassment if they do not perform to their expectations during the test scenarios, experience simulator sickness, or when physiological sensors are applied to their torso. To reduce the likelihood of embarrassment, you will perform the task individually and are reminded that your performance in the virtual environment will be anonymous. The research team will reassure you that the purpose of the study is not to assess your ability but to assess the technology.

Confidentiality and Anonymity:

The ethical duty of confidentiality includes safeguarding participants' identities, personal information, and data from unauthorized access, use, or disclosure.

Protecting your privacy and maintaining confidentiality is an important goal of the research team. Every effort to protect your privacy will be made. However it cannot be guaranteed. For example we may be required by law to allow access to research records.

When you sign this consent form you give us permission to

- Collect information from you
- Share information with the people conducting the study
- Share information with the people responsible for protecting your safety

The members of the research team will see study records that identify you by name. Other people may need to look at the study records that identify you by name. This might include the research ethics board. They can look at your records only when one of the research team is present.

Anonymity refers to protecting participants' identifying characteristics, such as name or description of physical appearance. Protecting your privacy and ensuring all personal data recorded during participation remains anonymous is an important goal for the research team. You will not be required to attend group session during this study. All participation will be conducted individually. Every reasonable effort will be made to assure your anonymity. You will also not be identified in any reports or publications.

Recording and Storage of Data:
The research team will collect and use only the information they need for this research study. This information will include your:

- date of birth
- gender
- performance metrics
- physiological data
- subjective assessments

Performance metrics will be recorded electronically during computer-based activities: time to complete, route selection and errors. Physiological parameters will be collected to assess stress experienced during the test trials: heart rate (EKG), galvanic skin response, respiration rate and skin temperature. Your response to subjective assessments like the SSQ and PTQ will also be reviewed and assessed.

Your name and contact information will be kept in a locked office on a password protected PC by the research team at MUN. It will not be shared with others without your permission. You will receive an alphanumeric participant code. All information collected from you will be recorded with the participant code and you will not be identifiable in the documentation and data. Your name will not appear in any report or article published as a result of this study

Information collected for this study will be kept for 5 years. Following this period, all electronic records of your participation will be permanently deleted and all paper files will be appropriately destroyed.

Reporting of Results:

The research team intends to publish the findings of this study in peer reviewed journals and academic conferences. Formal reports will be made available to the funding representatives. The data will be reported in a summarized statistical and descriptive form.

Sharing of Results with Participants:

On completion of data analysis, a report will be prepared for dissemination. Participants who wish to be informed of the results will have the opportunity to receive a copy of the final report.

ICEHR Statement:

The proposal for this research has been reviewed by the Interdisciplinary Committee on Ethics in Human Research and found to be in compliance with Memorial University's ethics policy. If you have ethical concerns about the research, such as the way you have been treated or your rights as a participant, you may contact the Chairperson of the ICEHR at icehr@mun.ca or by telephone at 709-864-2861.

Consent:

Your signature on this form means that:

- You have read the information about the research.
- You have been able to ask questions about this study.
- You are satisfied with the answers to all your questions.
- You understand what the study is about and what you will be doing.
- You understand that you are free to withdraw participation in the study without having to give a reason, and that doing so will not affect you now or in the future.
- You understand that if you choose to end participation **during** data collection, any data collected from you up to that point will be destroyed.
- You understand that if you choose to withdraw **after** data collection has ended, your data can be removed from the study up to two weeks after the completion of your participation.

☐ I agree to having all of the following physiological parameters recorded during my participation in this study.

☐ Heart Rate (EKG)
☐ Galvanic Skin Response

☐ Skin Temperature
☐ Respiration

☐ I agree to the use of my responses to all questionnaires completed during my participation in this study.

By signing this form, you do not give up your legal rights and do not release the researchers from their professional responsibilities.

Your signature confirms:

☐ I have read what this study is about and understood the risks and benefits. I have had adequate time to think about this and had the opportunity to ask questions and my questions have been answered.

☐ I agree to participate in the research project understanding the risks and contributions of my participation, that my participation is voluntary, and that I may end my participation.

☐ A copy of this Informed Consent Form has been given to me for my records.

Signature of participant Date

Researcher's Signature:
I have explained this study to the best of my ability. I invited questions and gave answers. I believe that the participant fully understands what is involved in being in the study, any potential risks of the study and that he or she has freely chosen to be in the study.

Signature of Principal Investigator Date

Appendix B: Performance evaluation rubric

The scoring rubric for the scenarios were based off the work presented in (Smith & Veitch, 2015).

Table 13. Performance rubric for scenario 1 and 2

Learning Objectives	**Specific Tasks**	**Performance Measure**	**Weighting**	
LO1. Establish Spatial Awareness of Environment	Identify Primary Muster Station	Correct location	See LO2	
LO2. Alarms Recognition: Understand role of alarms and urgency of situation	Identify General Platform Alarm (GPA)	Correct location (GPA = Mess Hall, PAPA = Lifeboat)	25	25
LO3. Routes and Mapping: Determine Primary and Alternative Routes to Muster Stations	Accommodation Cabin to Primary Muster Station	Route selected (15 points primary; 7.5 secondary; 0 lost or off route)	15	30
	Primary Muster Station back to Cabin	Route selected (15 points primary; 7.5 secondary; 0 lost or off route)	15	
LO5. Perform Muster Station Protocol and Individual Responsibilities	Perform T-Card Procedure at Muster Station	Correct location + Move t-card correctly	12.5	25
		Un-muster	12.5	

LO6. Safe Practices	Do not run on the platform	Speed of trainee (% running)	10	25
	Recognize and Use Fire Doors & Water Tight Doors	Number of fire/water tight doors left open (closed)	15	
LO7. First Actions - Taking PPE from Cabin	Know to locate and bring the following: Grab Bag and Immersion Suit	Takes Grab Bag and Immersion Suit	10	10
		Total	115	115

Table 14. Performance rubric for scenario 3

Learning Objectives	**Specific Tasks**	**Performance Measure**	**Weighting**	
LO1. Establish Spatial Awareness of Environment	Identify Primary Muster Station	Correct location	See LO2	
LO2. Alarms Recognition: Understand role of alarms and urgency of situation	Identify General Platform Alarm (GPA)	Correct location (GPA = Mess Hall, PAPA = Lifeboat)	25	25
LO3. Routes and Mapping: Determine Primary and Alternative Routes to Muster Stations	Accommodation Cabin to Primary Muster Station	Route selected (15 points primary; any other route = 0)	15	65
	Take safest route from primary Muster Station back to Cabin	Route selected (second) and re-route in event of alarm change/ PA update	10	
	Listen to PA and avoid blocked routes	Re-route in event of encounter hazard (most efficient route selected when re-routing)	10	
	Avoid Exposure to Hazards along path	Exposure to hazard = gas	15	
	Primary Muster Station back to Cabin	Route selected (prim, second,)	15	
LO5. Perform Muster Station Protocol and Individual Responsibilities	Perform T-Card Procedure at Muster Station	Correct location + Move t-card correctly	12.5	25
		Un-muster	12.5	

LO6. Safe Practices	Do not run on the platform Recognize and Use Fire Doors & Water Tight Doors	Speed of trainee (% running) Number of fire/water tight doors left open (closed)	10 15	25
LO7. First Actions - Taking PPE from Cabin	Know to locate and bring the following: Grab Bag and Immersion Suit	Takes Grab Bag and Immersion Suit	10	10
		Total	15 0	15 0

Table 15. Performance rubric for scenario 4 and 6

Learning Objectives	Specific Tasks	Performance Measure	Weighting	
LO1. Establish Spatial Awareness of Environment	Identify Primary Muster Station	Correct location	See LO2	
LO2. Alarms Recognition: Understand role of alarms and urgency of situation	Identify General Platform Alarm (GPA)	Correct location (GPA = Mess Hall)	25	50
	Identify PAPA	Correct location (PAPA = Lifeboat)	25	
LO3. Routes and Mapping: Determine Primary and Alternative Routes to Muster Stations	Accommodation CCR to Primary Muster Station	Route selected (15 points primary; 7.5 secondary; 0 lost or off route)	15	15
LO5. Perform Muster Station Protocol and Individual Responsibilities	Perform T-Card Procedure at Muster Station	Correct location + Move t-card correctly	12.5	25
		Transfer to Lifeboat Station and muster at lifeboat	12.5	
LO6. Safe Practices	Do not run on the platform	Speed of trainee (% running)	10	25
	Recognize and Use Fire Doors & Water Tight Doors	Number of fire/water tight doors left open (closed)	15	
LO7. First Actions - Taking PPE from Cabin	Know to locate and bring the following: Grab Bag and Immersion Suit	Takes Grab Bag and Immersion Suit	10	15
	Don Immersion Suit at Lifeboat Station	Put on Immersion Suit	5	
		Total	130	130

Table 16. Performance rubric for scenario 5

Learning Objectives	Specific Tasks	Performance Measure	Weighting	
LO1. Establish Spatial Awareness of Environment	Identify Primary Muster Station	Correct location	See LO2	
LO2. Alarms Recognition: Understand role of alarms and urgency of situation	Identify General Platform Alarm (GPA)	Correct location (GPA = Mess Hall, PAPA = Lifeboat)	25	25
LO3. Routes and Mapping: Determine Primary and Alternative Routes to Muster Stations	Accommodation Cabin to Primary Muster Station	Route selected (prime) & off route 15 points primary; any other route = 0	15	55
	Listen to PA and avoid blocked routes	Re-route in event of encounter hazard (most efficient route selected when re-routing)	10	
	Avoid Exposure to Hazards along path	Exposure to hazard = gas	15	
	Primary Muster Station back to Cabin	Route selected (prim, second,)	15	
LO5. Perform Muster Station Protocol and Individual Responsibilities	Perform T-Card Procedure at Muster Station	Correct location + Move t-card correctly	12.5	25
		Un-muster	12.5	
LO6. Safe Practices	Do not run on the platform	Speed of trainee (% running)	10	25
	Recognize and Use Fire Doors & Water Tight Doors	Number of fire/water tight doors left open (closed)	15	

LO7. First Actions - Taking PPE from Cabin	Know to locate and bring the following: Grab Bag and Immersion Suit	Takes Grab Bag and Immersion Suit	10	10
		Total	14 0	14 0

Table 17. Performance rubric for scenario 7

Learning Objectives	Specific Tasks	Performance Measure	Weighting	
LO1. Establish Spatial Awareness of Environment	Identify Primary Muster Station	Correct location	See LO2	
LO2. Alarms Recognition: Understand role of alarms and urgency of situation	Identify General Platform Alarm (GPA)	Correct location (GPA = Mess Hall)	25	50
	Identify PAPA	Correct location (PAPA = Lifeboat)	25	
LO3. Routes and Mapping: Determine <u>Primary and Alternative</u> Routes to Muster Stations	Accommodation CCR to Primary Muster Station	Route selected (second) & off route 15 points secondary; 0 all other routes	15	50
	Take safest route from primary Muster Station back to Cabin	Route selected (primary) and re-route in event of alarm change/ PA update	10	
	Listen to PA and avoid blocked routes	Re-route in event of encounter hazard (most efficient route selected when re-routing)	10	
	Avoid Exposure to Hazards along path	Exposure to hazard = gas	15	
LO5. Perform Muster Station Protocol and Individual Responsibilities	Perform T-Card Procedure at Muster Station	Correct location + Move t-card correctly	12.5	25
		Transfer to Lifeboat Station and muster at lifeboat	12.5	
LO6. Safe Practices	Do not run on the platform	Speed of trainee (% running)	10	25
	Recognize and Use Fire Doors & Water Tight Doors	Number of fire/water tight doors left open (closed)	15	
LO7. First Actions -	Know to locate and bring the following: Grab Bag and Immersion Suit	Takes Grab Bag and Immersion Suit	10	15

Taking PPE from Cabin	Don Immersion Suit at Lifeboat Station	Put on Immersion Suit	5	
		Total	16 5	16 5

Table 18. Performance rubric for scenario 8

Learning Objectives	Specific Tasks	Performance Measure	Weighting	
LO1. Establish Spatial Awareness of Environment	Identify Primary Muster Station	Correct location	See LO2	
LO2. Alarms Recognition: Understand role of alarms and urgency of situation	Identify General Platform Alarm (GPA)	Correct location (GPA = Mess Hall)	25	50
	Identify PAPA	Correct location (PAPA = Lifeboat)	25	
LO3. Routes and Mapping: Determine <u>Primary and Alternative</u> Routes to Muster Stations	Accommodation CCR to Primary Muster Station	Route selected (primary) & off route 15 points primary; 0 all other routes	15	40
	Listen to PA and avoid blocked routes	Re-route in event of encounter hazard (most efficient route selected when re-routing)	10	
	Avoid Exposure to Hazards along path	Exposure to hazard = gas	15	
LO5. Perform Muster Station Protocol and Individual Responsibilitie s	Perform T-Card Procedure at Muster Station	Correct location + Move t-card correctly	12 .5	25
		Transfer to Lifeboat Station and muster at lifeboat	12 .5	
	Do not run on the platform	Speed of trainee (% running)	10	25

LO6. Safe Practices	Recognize and Use Fire Doors & Water Tight Doors	Number of fire/water tight doors left open (closed)	15	
LO7. First Actions - Taking PPE from Cabin	Know to locate and bring the following: Grab Bag and Immersion Suit	Takes Grab Bag and Immersion Suit	10	15
	Don Immersion Suit at Lifeboat Station	Put on Immersion Suit	5	
		Total	15 5	15 5

Appendix C: Detailed results

C.1 Performance Scores data

Table 19. Performance scores results for scenario 1

No.	Alarm	Select Route (muster)	Select Route (return to cabin)	Register at TSR	Un-Register at TSR	Takes Equip	Not Running	Closing Doors	Total Score	%
1	25	7.5	15	12.50	12.50	10	10	15	107.5	93
2	25	7.5	15	12.50	12.50	10	10	15	107.5	93
3	25	15	15	12.50	12.50	10	10	15	115	100
4	25	15	15	12.50	12.50	10	10	0	100	87
5	25	15	15	12.50	12.50	10	10	15	115	100
6	25	15	15	12.50	12.50	10	10	15	115	100
7	25	15	15	12.50	12.50	10	10	15	115	100
8	25	15	15	12.50	12.50	10	10	15	115	100
9	25	7.5	15	12.50	12.50	10	10	15	107.5	93
10	25	15	15	12.50	12.50	10	10	15	115	100
11	25	15	15	12.50	12.50	10	10	15	115	100
12	25	15	15	12.50	12.50	10	10	15	115	100
13	25	15	15	12.50	12.50	10	10	15	115	100
14	25	15	15	12.50	12.50	10	10	15	115	100
15	25	15	0	12.50	12.50	10	10	15	100	87
16	25	15	15	12.50	12.50	10	10	15	115	100

17	25	15	15	12.50	12.50	10	10	15	115	100
18	25	15	0	12.50	12.50	10	10	15	100	87
19	25	7.5	15	12.50	12.50	10	10	15	107.5	93
20	25	15	15	12.50	12.50	10	10	15	115	100
21	25	15	15	12.50	12.50	10	10	15	115	100
22	25	15	15	12.50	12.50	10	10	15	115	100
23	25	15	15	12.50	12.50	10	10	15	115	100
24	25	15	15	12.50	12.50	10	10	15	115	100
25	25	15	15	12.50	12.50	10	10	15	115	100
26	25	7.5	15	12.50	12.50	10	10	15	107.5	93
27	25	15	15	12.50	12.50	10	10	15	115	100
28	25	7.5	15	12.50	12.50	10	10	15	107.5	93
29	25	15	15	12.50	12.50	10	10	15	115	100
30	25	7.5	15	12.50	12.50	10	10	15	107.5	93
31	25	15	15	12.50	12.50	10	10	15	115	100
32	25	7.5	15	12.50	12.50	10	10	15	107.5	93
33	25	15	15	12.50	12.50	10	10	15	115	100
34	25	15	15	12.50	12.50	10	10	15	115	100
35	25	15	15	12.50	12.50	10	10	15	115	100
36	25	15	15	12.50	12.50	10	10	15	115	100
37	25	15	15	12.50	12.50	10	10	15	115	100
38	25	15	15	12.50	12.50	10	10	15	115	100

Table 20. Performance scores results for scenario 2

No.	Alarm	Select Route (muster)	Select Route (return to cabin)	Register at TSR	Un-Register at TSR	Takes Equip	Not Running	Closing Doors	Total Score	%
1	25	15	15	12.50	12.50	10	10	15	115	100
2	25	15	15	12.50	12.50	10	10	15	115	100
3	25	15	15	12.50	12.50	10	10	15	115	100
4	25	15	15	12.50	12.50	10	10	15	115	100
5	25	15	15	12.50	12.50	10	10	15	115	100
6	25	15	15	12.50	12.50	10	10	15	115	100
7	25	15	15	12.50	12.50	10	10	15	115	100
8	25	15	15	12.50	12.50	10	10	15	115	100
9	25	15	15	12.50	12.50	10	10	15	115	100
10	25	15	15	12.50	12.50	10	10	15	115	100
11	25	15	15	12.50	12.50	10	10	15	115	100
12	25	15	15	12.50	12.50	10	10	15	115	100
13	25	15	15	12.50	12.50	10	10	15	115	100
14	25	15	15	12.50	12.50	10	10	15	115	100
15	25	15	15	12.50	12.50	10	10	15	115	100
16	25	15	15	12.50	12.50	10	10	15	115	100
17	25	15	15	12.50	12.50	10	10	15	115	100
18	25	15	15	12.50	12.50	10	10	15	115	100
19	25	15	15	12.50	12.50	10	10	15	115	100

20	25	15	15	12.50	12.50	10	10	15	115	100
21	25	15	15	12.50	12.50	10	10	15	115	100
22	25	15	15	12.50	12.50	10	10	15	115	100
23	25	15	15	12.50	12.50	10	10	15	115	100
24	25	15	15	12.50	12.50	10	10	15	115	100
25	25	15	15	12.50	12.50	10	10	15	115	100
26	25	15	15	12.50	12.50	10	10	15	115	100
27	25	15	15	12.50	12.50	10	10	15	115	100
28	25	15	15	12.50	12.50	10	10	15	115	100
29	25	15	15	12.50	12.50	10	10	15	115	100
30	25	15	15	12.50	12.50	10	10	15	115	100
31	25	15	15	12.50	12.50	10	10	0	100	87
32	25	15	15	12.50	12.50	10	10	15	115	100
33	0	15	15	12.50	0.00	10	10	15	77.5	67
34	25	15	15	12.50	12.50	10	10	15	115	100
35	25	15	15	12.50	12.50	10	10	15	115	100
36	25	15	15	12.50	12.50	10	10	15	115	100
37	25	7.5	15	12.50	12.50	10	10	15	107.5	93
38	25	15	15	12.50	12.50	10	10	15	115	100

Table 21. Performance scores for scenario 3

No.	Alarm	Select Route (muster)	Re-route alarm changes	Re-route encounter hazard	Avoid Hazard	Select Route (return to cabin)	Register at TSR	Un-Register at TSR	Takes Equip	Not Running	Closing Doors	Total Score	%
1	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
2	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
3	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
4	0	0	0	10	15	0	0.00	0.00	0	10	15	50	33
5	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
6	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
7	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
8	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
9	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
10	25	0	0	10	15	0	12.50	12.50	10	10	15	110	73
11	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
12	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
13	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
14	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
15	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
16	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
17	25	0	0	10	15	15	12.50	12.50	0	10	15	115	77
18	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100

19	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
20	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
21	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
22	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
23	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
24	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
25	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
26	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
27	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
28	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
29	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
30	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
31	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
32	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
33	0	0	0	0	0	0	12.50	0.00	10	10	15	47.5	32
34	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
35	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
36	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
37	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100
38	25	15	10	10	15	15	12.50	12.50	10	10	15	150	100

Table 22. Performance scores for scenario 4

No.	Alarm	Identify PAPA (Lifeboat muster)	Select Route (muster)	Register at TSR	Register at lifeboat	Put on Immersion suit	Takes Equip	Not Running	Closing Doors	Total Score	%
1	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
2	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
3	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
4	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
5	25	25	15	12.50	12.50	5	10	10	15	130	100
6	25	25	15	12.50	12.50	5	10	10	15	130	100
7	0	0	0	0.00	0.00	0	0	10	15	25	19
8	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
9	25	25	15	12.50	12.50	5	0	10	0	105	81
10	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
11	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
12	25	25	7.5	12.50	12.50	5	10	10	0	107.5	83
13	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
14	25	25	15	12.50	12.50	5	10	10	15	130	100
15	25	25	15	12.50	12.50	5	10	0	0	105	81
16	25	25	7.5	12.50	12.50	0	10	0	15	107.5	83
17	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
18	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94

19	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
20	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
21	25	25	7.5	12.50	12.50	0	0	10	15	107.5	83
22	25	25	15	12.50	12.50	5	10	10	15	130	100
23	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
24	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
25	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
26	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
27	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
28	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
29	25	25	7.5	12.50	12.50	0	0	10	15	107.5	83
30	25	25	7.5	12.50	12.50	0	0	10	15	107.5	83
31	25	25	7.5	12.50	12.50	5	0	10	15	112.5	87
32	0	0	0	0.00	0.00	0	0	10	15	25	19
33	25	25	15	12.50	12.50	0	0	10	15	115	88
34	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
35	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
36	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
37	25	25	15	12.50	12.50	5	10	10	15	130	100
38	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94

Table 23. Performance score for scenario 5

No.	Alarm	Select Route (muster)	Re-route when encounter hazard	Avoid Hazard	Select Route (return to cabin)	Register at TSR	Un-register at TSR	Takes Equip	Not Running	Closing Doors	Total Score	%
1	25	15	10	15	15	12.50	12.50	10	10	15	140	100
2	25	15	10	15	15	12.50	12.50	10	10	15	140	100
3	25	15	10	15	15	12.50	12.50	10	10	15	140	100
4	25	15	10	15	15	12.50	12.50	10	10	15	140	100
5	25	15	10	15	8	12.50	12.50	10	10	15	132.5	95
6	25	15	10	15	15	12.50	12.50	10	10	15	140	100
7	25	15	10	15	15	12.50	12.50	10	10	15	140	100
8	25	15	10	15	8	12.50	12.50	10	10	15	132.5	95
9	25	15	10	15	8	12.50	12.50	10	10	15	132.5	95
10	25	15	10	15	15	12.50	12.50	10	10	15	140	100
11	25	15	10	15	15	12.50	12.50	10	10	15	140	100
12	25	15	10	15	15	12.50	12.50	10	10	15	140	100
13	25	15	10	15	15	12.50	12.50	10	10	15	140	100
14	25	15	10	15	8	12.50	12.50	10	10	15	132.5	95
15	25	15	10	15	8	12.50	12.50	10	10	15	132.5	95
16	25	15	10	15	15	12.50	12.50	10	10	15	140	100
17	25	15	10	15	8	12.50	12.50	10	10	15	132.5	95
18	25	15	10	15	15	12.50	0.00	10	10	15	127.5	91

19	25	15	10	15	0	12.50	12.50	10	10	15	125	89
20	25	15	10	15	8	12.50	12.50	10	10	15	132.5	95
21	25	15	10	15	15	12.50	12.50	10	10	15	140	100
22	25	15	10	15	8	12.50	12.50	10	10	15	132.5	95
23	25	15	10	15	15	12.50	12.50	10	10	15	140	100
24	25	15	10	15	15	12.50	12.50	10	10	15	140	100
25	25	15	10	15	8	12.50	12.50	10	10	15	132.5	95
26	25	15	10	15	15	12.50	12.50	10	10	15	140	100
27	25	15	10	15	15	12.50	12.50	10	10	15	140	100
28	25	15	10	15	15	12.50	12.50	10	10	15	140	100
29	25	15	10	15	15	12.50	12.50	10	10	15	140	100
30	25	15	10	15	15	12.50	12.50	10	10	15	140	100
31	25	15	10	15	8	12.50	12.50	10	10	15	132.5	95
32	25	15	10	15	15	12.50	12.50	10	10	15	140	100
33	25	15	10	15	15	12.50	12.50	10	10	15	140	100
34	25	15	10	15	15	12.50	12.50	10	10	15	140	100
35	25	0	10	15	15	12.50	12.50	10	10	15	125	89
36	25	15	10	15	0	12.50	12.50	10	10	15	125	89
37	25	15	10	15	15	12.50	12.50	10	10	15	140	100
38	25	15	10	15	8	12.50	12.50	10	10	15	132.5	95

Table 24. Performance scores for scenario 6

No.	Alarm	Identify PAPA (Lifeboat muster)	Route Selected (muster)	Register at TSR	Register at Lifeboat	Put on Immersion suit	Takes Equip	Not Running	Closing Doors	Total Score	%
1	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
2	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
3	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
4	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
5	25	25	15	12.50	12.50	5	10	10	15	130	100
6	25	25	15	12.50	12.50	5	10	0	15	120	92
7	25	25	7.5	12.50	12.50	5	0	10	15	112.5	87
8	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
9	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
10	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
11	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
12	25	25	7.5	12.50	12.50	5	10	10	0	107.5	83
13	25	25	15	12.50	12.50	5	10	10	15	130	100
14	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
15	25	25	15	12.50	12.50	5	10	10	15	130	100
16	0	0	0	0.00	0.00	0	0	10	15	25	19
17	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
18	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94

19	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
20	25	25	7.5	12.50	12.50	5	10	10	0	107.5	83
21	25	25	7.5	12.50	12.50	5	0	10	15	112.5	87
22	25	25	15	12.50	12.50	5	10	10	15	130	100
23	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
24	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
25	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
26	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
27	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
28	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
29	25	25	7.5	12.50	12.50	5	0	10	15	112.5	87
30	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
31	25	25	7.5	12.50	12.50	5	0	10	15	112.5	87
32	0	0	0	0.00	0.00	0	0	10	15	25	19
33	25	25	15	12.50	12.50	5	10	10	15	130	100
34	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
35	25	0	7.5	12.50	0.00	0	0	10	15	70	54
36	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94
37	25	25	7.5	12.50	12.50	0	10	10	15	117.5	90
38	25	25	7.5	12.50	12.50	5	10	10	15	122.5	94

Table 25. Performance scores for scenario 7

No.	Alarm	Identify PAPA (Lifeboat muster)	Select Route (muster)	Re-Route alarm changes	Re-route encounter hazard	Avoids Hazard	Register at TSR	Register at lifeboat	Put on immersion suit	Takes Equip	Not Running	Closing Doors	Total Score	%
1	25	25	15	0	10	15	12.50	12.50	5	10	10	15	155	94
2	25	25	15	10	10	15	12.50	12.50	5	10	10	15	165	100
3	25	25	15	10	10	15	12.50	12.50	5	10	10	15	165	100
4	25	25	15	10	10	15	12.50	12.50	5	10	10	15	165	100
5	25	25	15	10	10	15	12.50	12.50	5	10	10	15	165	100
6	25	25	15	10	10	15	12.50	12.50	5	10	10	15	165	100
7	25	25	0	10	10	15	12.50	12.50	5	0	10	15	140	85
8	25	25	15	10	10	15	12.50	12.50	5	10	10	15	165	100
9	25	25	15	10	10	15	12.50	12.50	5	10	10	15	165	100
10	25	25	15	10	10	15	12.50	12.50	5	10	10	0	150	91
11	25	25	15	10	10	15	12.50	12.50	5	10	10	15	165	100
12	25	25	15	10	10	15	12.50	12.50	5	10	10	0	150	91
13	25	25	0	0	10	15	12.50	12.50	5	10	10	15	140	85
14	25	25	15	10	10	15	12.50	12.50	5	10	10	15	165	100
15	25	25	0	0	10	15	12.50	12.50	5	10	10	15	140	85
16	0	0	0	0	10	15	0.00	0.00	0	0	10	15	50	30
17	25	25	15	10	10	15	12.50	12.50	5	10	10	15	165	100

18	25	25	15	10	10	15	12.50	12.50	5	10	10	0	150	91
19	25	25	15	10	10	15	12.50	12.50	5	10	10	15	165	100
20	25	25	15	10	10	15	12.50	12.50	5	10	10	15	165	100
21	25	25	0	10	10	15	12.50	12.50	5	0	10	15	140	85
22	25	25	0	0	10	15	12.50	12.50	5	10	10	15	140	85
23	25	25	15	10	10	15	12.50	12.50	5	10	10	15	165	100
24	25	25	15	10	10	15	12.50	12.50	5	10	10	15	165	100
25	25	25	15	10	10	15	12.50	12.50	5	10	10	15	165	100
26	25	25	15	10	10	15	12.50	12.50	5	10	10	15	165	100
27	25	25	15	10	10	15	12.50	12.50	5	10	10	15	165	100
28	25	25	15	10	10	15	12.50	12.50	5	10	10	15	165	100
29	25	25	0	10	10	15	12.50	12.50	5	0	10	15	140	85
30	25	25	15	10	10	15	12.50	12.50	5	10	10	15	165	100
31	25	25	0	10	10	15	12.50	12.50	5	0	10	15	140	85
32	25	25	15	10	10	15	12.50	12.50	5	10	10	15	165	100
33	25	25	0	0	10	15	12.50	12.50	5	10	10	15	140	85
34	25	25	15	10	10	15	12.50	12.50	5	10	10	15	165	100
35	25	25	15	10	10	15	12.50	12.50	5	10	10	15	165	100
36	25	25	15	10	10	15	12.50	12.50	5	10	10	15	165	100
37	25	25	0	0	10	15	12.50	12.50	5	10	10	15	140	85
38	25	25	15	10	10	15	12.50	12.50	5	10	10	15	165	100

Table 26. Performance scores for scenario 8

No.	Alarm	Identify PAPA (Life boat muster)	Select Route (muster)	Re-route when encounter hazard	Avoids Hazard	Register at TSR	Register at lifeboat	Put on immersion suit	Takes Equip	Not Running	Closing Doors	Total Score	%
1	25	25	0	10	15	12.50	12.50	5	10	10	15	140	90
2	25	25	0	10	15	12.50	12.50	5	10	10	0	125	81
3	25	25	0	10	15	12.50	12.50	5	10	10	15	140	90
4	0	0	0	0	15	0.00	0.00	0	0	10	0	25	16
5	25	25	0	10	15	12.50	12.50	5	10	10	15	140	90
6	25	25	15	10	15	12.50	12.50	5	10	10	15	155	100
7	25	25	0	10	15	12.50	12.50	5	0	10	15	130	84
8	25	25	0	10	15	12.50	12.50	5	10	10	15	140	90
9	25	25	0	10	15	12.50	12.50	5	0	10	15	130	84
10	25	25	0	10	15	12.50	12.50	5	10	10	15	140	90
11	0	0	0	0	15	0.00	0.00	0	0	10	15	40	26
12	25	25	0	10	15	12.50	12.50	5	10	10	0	125	81
13	25	25	0	10	15	12.50	12.50	5	10	10	15	140	90
14	25	25	15	10	15	12.50	12.50	5	10	10	15	155	100
15	25	25	15	10	15	12.50	12.50	5	10	10	15	155	100
16	0	0	0	10	15	0.00	0.00	0	0	10	15	50	32
17	25	25	0	10	15	12.50	12.50	5	10	10	15	140	90

18	25	25	0	10	15	12.50	12.50	5	10	10	15	140	90
19	0	0	0	0	0	0.00	0.00	0	0	10	15	25	16
20	25	25	0	10	15	12.50	12.50	5	10	10	15	140	90
21	25	25	0	10	15	12.50	12.50	5	0	10	15	130	84
22	25	25	15	10	15	12.50	12.50	5	10	10	15	155	100
23	25	25	0	10	15	12.50	12.50	5	10	10	15	140	90
24	25	25	0	10	15	12.50	12.50	5	10	10	15	140	90
25	25	25	0	10	15	12.50	12.50	5	10	10	15	140	90
26	25	25	0	10	15	12.50	12.50	5	10	10	15	140	90
27	25	25	0	10	15	12.50	12.50	5	10	10	15	140	90
28	25	25	15	10	15	12.50	12.50	5	10	10	15	155	100
29	25	25	0	10	15	12.50	12.50	5	0	10	15	130	84
30	25	25	0	0	0	12.50	12.50	5	10	10	15	115	74
31	25	25	0	10	15	12.50	12.50	5	0	10	15	130	84
32	25	25	0	10	15	12.50	12.50	5	10	10	15	140	90
33	25	25	15	10	15	12.50	12.50	5	10	10	15	155	100
34	25	25	0	10	15	12.50	12.50	5	10	10	15	140	90
35	25	25	0	10	15	12.50	12.50	0	0	10	15	125	81
36	0	0	7.5	10	15	0.00	0.00	0	0	10	15	57.5	37
37	25	25	15	10	15	12.50	12.50	5	10	10	15	155	100
38	25	25	0	10	15	12.50	12.50	5	10	10	15	140	90

C.2 Physiological measures

C.2.1 Classification results

Table 27. Classification with LDA, with greedy search algorithm using to select features

Subject No.	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	Scenario 7	Scenario 8
1	78.8	56.7	77.0	75.1	63.7	76.5	88.8	97.2
2	98.9	89.6	87.8	75.8	99.8	88.8	87.1	93.0
3	78.1	69.6	68.6	64.2	65.5	74.5	75.8	74.6
4	82.6	63.4	87.0	91.9	85.7	87.5	81.8	87.6
5	86.9	74.7	82.8	74.2	76.0	82.8	79.0	81.7
6	81.2	89.2	89.2	87.7	89.2	78.7	93.3	86.1
7	87.7	83.1	80.4	75.4	65.2	84.1	73.9	84.3
8	87.9	94.9	83.9	88.1	96.0	70.7	77.4	84.1
9	92.7	89.7	77.0	90.1	75.9	77.3	64.2	63.0
10	92.9	83.6	75.8	94.4	67.9	84.7	85.6	76.2
11	86.0	87.7	83.7	79.8	83.4	84.0	85.1	71.1
12	83.0	75.0	69.2	88.6	57.8	96.7	82.3	87.4
13	69.7	70.3	93.3	95.3	86.6	73.0	91.1	80.7
14	73.9	84.3	83.8	79.0	85.2	72.4	92.8	78.2
15	96.7	98.3	99.1	100.0	95.8	98.1	99.8	99.7
16	80.4	61.5	68.1	73.3	74.6	70.4	71.4	72.7

17	88.1	79.5	70.4	94.1	74.5	88.1	91.4	73.8
18	90.9	98.9	88.1	87.6	89.1	88.6	89.8	79.8
19	97.6	93.7	93.1	100.0	98.2	99.1	99.7	100.0
20	62.5	78.6	65.1	63.6	60.9	60.5	82.0	80.3
21	79.9	79.4	81.1	91.6	70.5	84.0	89.7	88.6
22	79.3	84.6	61.3	79.7	63.0	74.0	72.2	65.7
23	79.6	77.2	66.6	76.7	88.2	78.3	90.2	83.2
24	66.7	82.7	81.8	84.3	71.7	71.5	75.5	68.6
25	68.4	87.2	71.6	77.7	68.8	86.4	72.8	68.5
26	83.0	73.0	84.6	70.9	79.9	69.8	79.8	73.8
27	91.0	65.6	61.2	72.5	70.6	88.9	72.4	99.4
28	78.2	72.2	75.2	99.2	96.8	71.2	66.3	36.8
29	84.4	85.6	90.9	81.3	66.3	80.6	82.9	86.6
30	81.2	71.3	89.8	76.8	81.0	76.0	86.5	84.7
31	71.3	76.5	85.5	69.6	67.2	75.5	64.8	86.5
32	84.3	75.1	77.3	95.0	94.9	82.3	74.7	72.3
33	81.9	83.9	82.2	91.4	58.8	92.2	79.5	75.8
34	82.8	94.4	84.9	95.2	98.7	95.1	90.9	98.3
35	75.8	70.5	93.0	69.8	67.5	84.1	94.2	76.5
36	90.4	92.4	93.5	82.2	76.8	91.9	82.7	88.0
37	87.3	67.6	78.8	77.9	68.7	83.6	75.8	78.5
38	92.6	86.3	91.7	85.7	94.8	90.2	86.2	87.5

Average	**83.02**	**80.20**	**80.91**	**83.04**	**78.29**	**81.89**	**82.35**	**80.80**

Table 28. Classification with SVM, with greedy search algorithm using to select features

Subject No.	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	Scenario 7	Scenario 8
1	50.7	50.0	57.9	54.7	50.2	61.5	73.3	96.9
2	99.4	60.8	79.7	62.7	99.6	82.7	81.8	90.5
3	80.7	50.0	53.0	49.5	63.9	68.6	65.3	78.7
4	82.3	56.2	85.8	91.3	85.3	84.3	84.0	89.1
5	62.2	49.9	54.9	70.6	59.9	74.3	81.8	77.3
6	76.8	87.1	86.0	78.5	85.9	79.4	93.7	84.7
7	80.0	54.8	76.6	76.3	56.4	83.9	73.6	85.4
8	73.6	84.5	77.9	89.4	96.7	58.1	65.5	61.1
9	69.5	96.1	61.3	91.8	76.8	68.4	55.9	50.3
10	74.3	72.2	63.7	83.7	62.8	75.5	80.0	70.0
11	84.2	83.2	84.1	81.6	85.4	85.2	84.6	70.3
12	78.7	61.9	50.0	88.9	50.0	98.6	84.3	85.5
13	51.9	50.0	94.0	94.8	87.7	66.4	90.4	74.5
14	52.5	86.6	79.1	82.7	73.9	70.8	96.7	69.2
15	96.7	99.1	98.9	100.0	95.7	98.5	99.8	88.5
16	82.1	50.0	68.5	75.2	69.5	68.2	68.8	70.3
17	85.3	64.7	61.3	94.2	73.5	80.0	87.1	71.8

18	75.0	97.2	79.9	81.9	67.2	88.7	89.1	71.0
19	98.2	94.7	93.7	100.0	98.2	99.1	100.0	100.0
20	49.9	50.6	50.0	52.3	54.6	50.8	81.0	78.3
21	69.7	61.8	59.9	92.2	61.2	85.6	90.6	90.0
22	54.4	73.0	50.0	78.4	53.1	64.2	60.3	50.0
23	60.8	74.5	50.0	75.4	88.0	78.6	91.8	85.6
24	58.3	76.8	65.9	83.6	61.0	52.8	69.0	59.3
25	64.2	87.1	52.6	52.2	49.6	85.9	57.1	67.3
26	84.3	50.0	70.9	56.7	55.3	59.8	57.2	68.9
27	61.5	50.1	49.6	56.4	56.6	83.1	68.5	99.0
28	50.7	56.1	57.0	99.6	98.3	68.7	54.2	37.7
29	88.2	78.4	80.3	75.8	54.4	79.4	84.0	78.1
30	80.5	50.6	91.3	75.5	79.8	72.3	87.9	85.9
31	50.6	50.2	88.7	67.2	59.3	66.2	62.6	72.4
32	79.1	53.0	52.0	93.0	97.2	83.5	67.2	66.2
33	54.3	84.2	85.0	92.4	58.0	93.6	79.3	76.7
34	78.4	95.3	85.3	94.8	98.7	94.8	88.0	98.6
35	67.7	52.4	91.7	67.0	57.7	86.8	73.1	75.3
36	88.5	91.6	89.2	81.4	76.8	91.2	85.0	87.5
37	67.9	50.1	50.0	66.1	56.5	83.6	80.0	73.1
38	87.3	80.6	92.6	85.6	94.8	91.0	86.2	85.4
Average	**72.37**	**68.82**	**71.54**	**78.76**	**72.36**	**78.01**	**78.38**	**76.85**

Table 29. Classification by LDA, with full features

Subject No.	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	Scenario 7	Scenario 8
1	81.17	68.11	79.82	82.43	71.84	78.56	92.09	94.67
2	97.77	94.62	95.96	91.32	99.01	92.99	95.61	95.34
3	80.92	70.80	67.04	62.40	72.37	78.39	75.82	77.85
4	87.11	76.91	90.58	94.05	89.35	87.53	88.25	95.33
5	84.63	76.79	78.52	81.21	85.07	85.46	87.98	86.32
6	81.24	88.81	90.84	88.93	89.03	79.58	91.45	82.98
7	89.60	77.97	87.16	84.12	70.20	86.11	74.72	86.89
8	90.61	95.28	83.00	88.45	98.47	77.88	77.81	83.88
9	87.20	88.53	72.82	90.56	72.45	80.92	70.60	70.87
10	88.69	76.10	80.83	95.34	72.16	86.08	81.78	80.32
11	88.15	82.61	90.47	82.67	86.28	87.69	88.52	76.03
12	87.24	72.15	70.39	89.31	60.88	95.57	85.70	86.08
13	67.97	72.87	89.95	91.00	86.45	81.07	93.59	89.04
14	69.77	87.04	78.37	81.45	84.15	79.85	92.61	74.94
15	97.39	98.05	97.87	99.78	97.72	98.85	98.47	98.95
16	88.88	73.20	69.55	78.70	87.35	76.47	76.70	79.18
17	94.07	86.65	73.16	99.05	79.73	87.40	96.90	81.12
18	89.08	98.24	82.12	86.06	83.78	96.07	91.31	84.97
19	96.87	95.10	96.84	99.41	98.60	98.39	98.73	100.00
20	62.02	75.84	67.08	63.51	63.02	68.31	83.44	85.25

21	81.54	84.08	84.10	94.20	82.87	92.05	88.35	85.88
22	89.33	81.79	67.21	87.13	70.69	66.01	79.03	70.93
23	79.91	86.56	68.09	77.28	88.76	80.25	93.55	82.45
24	79.55	87.59	83.59	84.85	74.76	82.62	75.79	85.80
25	66.31	86.50	78.62	77.27	61.38	86.92	71.05	74.59
26	84.75	76.83	86.56	76.08	81.48	75.72	82.79	84.88
27	86.20	71.25	71.53	79.64	72.42	93.13	74.65	99.02
28	80.74	79.10	83.43	96.18	94.62	76.88	83.21	24.17
29	88.83	85.86	85.46	86.87	73.94	80.25	83.85	84.14
30	83.90	71.64	89.69	81.48	85.76	90.66	83.58	93.38
31	84.17	80.39	88.55	80.84	77.38	74.19	67.54	81.07
32	94.02	86.57	89.17	95.02	93.22	87.10	74.59	74.17
33	87.21	84.06	83.18	93.47	73.17	83.70	82.62	79.70
34	83.89	92.76	85.71	92.93	97.15	94.96	93.67	97.71
35	80.27	83.25	97.81	77.72	68.38	88.33	95.18	87.33
36	96.49	88.78	87.24	82.15	86.87	94.94	84.67	92.18
37	85.66	74.65	79.00	83.25	70.97	91.71	80.88	84.08
38	92.76	86.86	95.89	83.01	93.82	95.28	89.11	84.51
Average	**85.16**	**82.74**	**82.82**	**85.77**	**81.46**	**85.21**	**84.90**	**83.58**

Table 30. Classification by SVM, with full features

Subject No.	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	Scenario 7	Scenario 8
1	75.04	50.00	67.67	78.19	53.56	68.69	84.60	85.62
2	97.21	72.04	89.76	79.11	96.46	84.84	87.87	90.81
3	78.89	55.11	58.81	51.85	61.17	68.52	73.92	74.88
4	85.45	70.63	89.22	90.38	90.52	92.44	86.15	90.98
5	74.58	78.97	72.60	80.99	77.10	75.03	88.29	84.58
6	69.58	87.22	89.52	84.10	85.18	81.21	93.38	85.81
7	75.23	68.53	82.01	82.23	67.91	82.79	73.60	82.52
8	82.44	89.44	80.23	90.33	97.96	75.86	67.49	80.23
9	69.76	87.32	74.79	91.79	74.45	75.30	62.98	75.65
10	78.05	72.08	59.90	89.81	69.30	88.35	74.23	76.90
11	88.00	81.76	87.19	87.66	85.08	91.21	89.02	78.82
12	85.31	80.07	59.50	87.44	54.00	95.23	85.93	87.37
13	60.93	58.30	89.81	91.30	80.43	71.82	88.74	86.04
14	55.43	83.08	77.01	81.51	83.41	78.35	92.51	71.40
15	97.23	92.10	95.82	98.65	95.23	97.57	97.34	95.40
16	78.57	59.93	72.37	84.37	82.62	73.82	78.85	77.42
17	87.40	66.31	70.40	97.10	69.72	76.97	93.49	76.01
18	83.31	97.23	74.63	87.13	75.73	93.22	87.09	82.45
19	98.66	96.07	96.12	99.50	97.42	98.38	99.42	99.93
20	55.31	69.33	56.40	73.32	65.21	67.77	85.22	84.64

21	74.03	63.73	69.52	94.08	73.34	93.58	86.38	87.07
22	56.98	80.19	49.79	76.65	63.29	64.04	67.85	50.93
23	72.14	80.18	60.95	81.01	85.49	82.59	92.41	83.78
24	58.02	82.49	69.74	82.35	76.28	68.75	75.88	68.00
25	64.87	79.99	71.91	59.90	61.49	84.52	65.95	70.54
26	80.24	68.69	82.86	68.33	66.79	73.31	62.14	84.47
27	70.74	62.93	51.07	63.46	64.52	86.14	69.32	93.57
28	68.81	62.52	62.36	95.66	94.87	70.56	63.84	22.79
29	79.52	81.77	77.57	76.93	72.93	82.47	85.07	75.45
30	83.05	69.12	91.62	82.85	83.95	86.60	79.75	92.38
31	67.65	70.10	88.97	77.32	73.06	67.16	71.70	81.66
32	87.60	61.60	79.18	96.41	89.56	86.92	74.35	76.03
33	84.78	85.03	90.24	92.50	64.08	91.03	86.26	85.48
34	76.64	97.09	78.80	89.50	98.48	94.98	93.53	98.84
35	62.70	74.83	94.82	67.53	65.38	87.90	85.59	82.68
36	94.82	90.68	89.78	83.96	85.53	91.92	84.31	93.82
37	84.64	69.81	75.53	78.74	67.94	93.17	83.11	84.35
38	84.95	81.45	87.05	84.77	93.05	88.88	86.36	84.57
Average	**77.07**	**75.73**	**76.72**	**83.12**	**77.43**	**82.42**	**81.68**	**81.15**

C.2.2 Clustering results

K-means clustering

Case 1: Cluster all data (scenario and baseline combined) of each participant

Table 31. Number of data points in each scenario which fell into cluster 1

No.	Sc. 1	Sc. 2	Sc. 3	Sc. 4	Sc. 5	Sc. 6	Sc. 7	Sc. 8	Baseline
1	11	3	14	4	16	16	18	18	231
2	19	11	12	30	8	21	21	0	353
3	10	5	14	12	5	53	28	7	228
4	9	17	4	17	9	17	33	10	331
5	5	8	12	4	6	10	10	10	107
6	7	0	10	9	11	7	3	11	69
7	13	7	6	8	12	34	18	13	216
8	2	20	26	35	23	22	42	49	199
9	15	14	10	15	20	9	7	19	107
10	10	18	20	10	19	38	19	11	307
11	1	0	4	2	5	9	5	8	298
12	9	7	9	22	19	12	10	15	162
13	4	6	1	24	8	6	3	29	206
14	19	23	14	21	11	12	23	27	236
15	5	1	2	1	4	1	0	55	375
16	2	52	16	9	16	102	21	0	276

17	4	4	4	71	37	18	9	3	272
18	16	26	6	16	18	34	27	12	111
19	60	37	15	70	9	20	5	37	16
20	2	2	2	3	3	2	4	3	84
21	11	19	4	54	15	13	5	6	264
22	7	1	8	6	9	6	7	5	153
23	15	6	8	17	18	29	5	0	245
24	14	14	17	3	37	16	20	13	199
25	26	10	19	24	14	25	13	23	218
26	2	3	2	0	1	10	13	5	66
27	13	23	24	29	18	31	19	11	277
28	14	10	12	2	10	23	4	22	153
29	4	24	24	43	6	4	6	28	250
30	20	4	23	9	7	11	8	10	111
31	7	3	1	30	17	8	51	4	240
32	28	14	28	10	23	81	20	34	249
33	24	0	43	0	0	13	46	15	206
34	28	9	5	5	0	0	1	0	413
35	18	22	3	11	26	16	30	44	398
36	54	0	0	0	0	0	0	0	111
37	25	11	25	28	34	24	40	35	228
38	31	23	32	49	30	42	38	42	147

Table 32. Proportion of data in each scenario which fell into cluster 1

No.	Sc. 1	Sc. 2	Sc. 3	Sc. 4	Sc. 5	Sc. 6	Sc. 7	Sc. 8	Baseline
1	23.40	9.09	38.89	8.16	35.56	32.65	28.13	34.62	28.70
2	34.55	28.21	28.57	43.48	14.55	36.21	36.21	0.00	44.13
3	17.24	11.11	30.43	20.69	9.09	58.89	50.00	10.00	28.29
4	12.86	28.81	3.92	18.28	10.23	18.48	37.93	8.77	40.42
5	12.50	20.51	30.00	8.33	10.34	21.28	15.63	16.95	13.28
6	14.58	0.00	14.08	12.86	15.49	10.29	2.91	17.19	8.65
7	30.95	17.95	12.50	7.69	20.34	51.52	27.27	20.63	26.83
8	4.44	43.48	52.00	43.21	29.87	27.85	66.67	73.13	24.60
9	27.27	33.33	23.26	23.08	27.40	11.39	9.72	30.16	13.14
10	28.57	48.65	37.74	20.00	38.78	69.09	38.78	22.92	37.71
11	2.00	0.00	8.00	2.38	7.58	10.71	5.95	7.77	37.02
12	23.68	18.42	21.43	32.84	35.85	23.53	19.61	28.30	20.05
13	9.52	15.79	2.56	40.68	14.81	12.00	4.62	44.62	25.75
14	46.34	56.10	34.15	25.30	15.94	18.46	40.35	47.37	27.67
15	6.58	1.85	3.33	1.04	4.49	1.52	0.00	77.46	46.53
16	3.92	98.11	27.59	8.65	25.00	98.08	20.19	0.00	34.37
17	7.55	8.16	5.63	69.61	46.25	25.00	18.37	3.41	33.83
18	27.59	56.52	13.64	25.40	31.58	51.52	45.00	20.00	13.67
19	78.95	60.66	23.44	80.46	8.33	20.41	4.95	75.51	2.00
20	4.35	4.55	4.55	4.29	4.00	2.70	5.48	3.33	10.50

21	25.00	46.34	9.52	52.94	27.27	17.11	7.81	11.32	32.59
22	17.07	2.50	21.62	13.64	14.52	11.11	9.72	12.20	19.13
23	34.88	14.29	19.05	28.81	31.58	33.72	6.49	0.00	30.28
24	35.90	37.84	42.50	5.45	46.84	29.63	35.71	24.53	24.75
25	33.77	26.32	47.50	46.15	22.95	46.30	25.00	24.73	26.72
26	4.08	8.33	5.41	0.00	2.04	17.86	26.00	9.62	8.24
27	31.71	54.76	55.81	47.54	34.62	40.79	27.14	21.15	34.45
28	27.45	27.03	28.57	3.08	18.87	39.66	7.27	17.60	18.19
29	7.69	46.15	53.33	42.16	10.71	4.40	10.00	45.90	30.56
30	27.78	8.70	41.07	8.82	12.28	15.07	11.76	13.33	13.84
31	15.22	6.82	2.00	43.48	25.37	14.55	44.74	6.90	29.93
32	46.67	33.33	54.90	10.64	38.98	79.41	30.77	34.34	30.97
33	50.00	0.00	69.35	0.00	0.00	19.40	52.27	17.65	25.62
34	35.44	12.00	9.80	7.58	0.00	0.00	1.52	0.00	50.43
35	40.91	47.83	6.52	16.18	48.15	15.69	48.39	43.14	49.44
36	100.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	13.67
37	55.56	20.00	56.82	57.14	57.63	33.80	50.00	66.04	28.11
38	64.58	57.50	71.11	68.06	41.10	72.41	63.33	63.64	18.24
Avg	**28.17**	**26.61**	**26.59**	**24.95**	**22.06**	**28.75**	**24.62**	**25.11**	**26.38**

Table 33. Number of data points in each scenario which fell into cluster 2

No.	Sc. 1	Sc. 2	Sc. 3	Sc. 4	Sc. 5	Sc. 6	Sc. 7	Sc. 8	Baseline
1	8	18	3	21	19	16	23	21	195
2	0	22	23	0	0	23	3	0	272
3	0	2	2	10	3	5	5	2	62
4	2	8	32	10	14	12	4	2	65
5	10	13	6	15	31	22	28	18	314
6	13	42	37	14	32	42	82	29	60
7	2	3	4	24	4	6	6	8	73
8	21	3	7	13	6	13	9	10	321
9	15	17	23	35	43	23	27	25	300
10	0	8	3	6	13	9	6	2	185
11	21	24	15	23	14	21	23	11	180
12	28	25	12	32	12	35	30	35	181
13	10	7	10	10	15	7	8	6	134
14	10	1	1	14	2	5	3	11	255
15	6	3	0	1	6	1	1	5	148
16	2	0	1	0	0	0	10	102	116
17	1	8	1	3	8	10	10	11	48
18	4	0	16	2	3	9	0	4	282
19	1	0	0	0	1	0	0	0	442
20	27	24	10	32	4	13	33	30	241

21	6	5	2	3	4	2	6	2	169
22	4	8	9	10	9	19	29	15	202
23	3	5	6	3	4	13	15	7	62
24	6	10	2	10	8	16	11	10	197
25	22	6	2	4	14	2	7	21	216
26	5	4	0	3	0	0	1	5	170
27	5	8	9	16	17	10	29	15	137
28	1	3	5	55	36	2	32	8	217
29	4	6	4	40	12	15	9	8	129
30	28	11	15	12	12	15	31	30	283
31	29	24	37	20	36	28	17	16	273
32	14	11	3	2	19	6	15	27	180
33	1	18	15	14	8	24	8	2	119
34	27	60	30	46	10	1	26	9	239
35	6	1	35	20	1	23	2	6	138
36	0	6	3	13	24	7	16	90	215
37	8	3	5	5	9	27	1	2	186
38	4	7	8	12	35	6	10	15	168

Table 34. Proportion of data in each scenario which fell into cluster 2

No.	Sc. 1	Sc. 2	Sc. 3	Sc. 4	Sc. 5	Sc. 6	Sc. 7	Sc. 8	Baseline
1	17.02	54.55	8.33	42.86	42.22	32.65	35.94	40.38	24.22
2	0.00	56.41	54.76	0.00	0.00	39.66	5.17	0.00	34.00
3	0.00	4.44	4.35	17.24	5.45	5.56	8.93	2.86	7.69
4	2.86	13.56	31.37	10.75	15.91	13.04	4.60	1.75	7.94
5	25.00	33.33	15.00	31.25	53.45	46.81	43.75	30.51	38.96
6	27.08	76.36	52.11	20.00	45.07	61.76	79.61	45.31	7.52
7	4.76	7.69	8.33	23.08	6.78	9.09	9.09	12.70	9.07
8	46.67	6.52	14.00	16.05	7.79	16.46	14.29	14.93	39.68
9	27.27	40.48	53.49	53.85	58.90	29.11	37.50	39.68	36.86
10	0.00	21.62	5.66	12.00	26.53	16.36	12.24	4.17	22.73
11	42.00	50.00	30.00	27.38	21.21	25.00	27.38	10.68	22.36
12	73.68	65.79	28.57	47.76	22.64	68.63	58.82	66.04	22.40
13	23.81	18.42	25.64	16.95	27.78	14.00	12.31	9.23	16.75
14	24.39	2.44	2.44	16.87	2.90	7.69	5.26	19.30	29.89
15	7.89	5.56	0.00	1.04	6.74	1.52	1.11	7.04	18.36
16	3.92	0.00	1.72	0.00	0.00	0.00	9.62	99.03	14.45
17	1.89	16.33	1.41	2.94	10.00	13.89	20.41	12.50	5.97
18	6.90	0.00	36.36	3.17	5.26	13.64	0.00	6.67	34.73
19	1.32	0.00	0.00	0.00	0.93	0.00	0.00	0.00	55.25
20	58.70	54.55	22.73	45.71	5.33	17.57	45.21	33.33	30.13

21	13.64	12.20	4.76	2.94	7.27	2.63	9.38	3.77	20.86
22	9.76	20.00	24.32	22.73	14.52	35.19	40.28	36.59	25.25
23	6.98	11.90	14.29	5.08	7.02	15.12	19.48	8.43	7.66
24	15.38	27.03	5.00	18.18	10.13	29.63	19.64	18.87	24.50
25	28.57	15.79	5.00	7.69	22.95	3.70	13.46	22.58	26.47
26	10.20	11.11	0.00	5.77	0.00	0.00	2.00	9.62	21.22
27	12.20	19.05	20.93	26.23	32.69	13.16	41.43	28.85	17.04
28	1.96	8.11	11.90	84.62	67.92	3.45	58.18	6.40	25.80
29	7.69	11.54	8.89	39.22	21.43	16.48	15.00	13.11	15.77
30	38.89	23.91	26.79	11.76	21.05	20.55	45.59	40.00	35.29
31	63.04	54.55	74.00	28.99	53.73	50.91	14.91	27.59	34.04
32	23.33	26.19	5.88	2.13	32.20	5.88	23.08	27.27	22.39
33	2.08	38.30	24.19	13.73	9.20	35.82	9.09	2.35	14.80
34	34.18	80.00	58.82	69.70	15.38	1.22	39.39	10.59	29.18
35	13.64	2.17	76.09	29.41	1.85	22.55	3.23	5.88	17.14
36	0.00	12.24	6.25	18.84	24.00	9.21	23.19	88.24	26.48
37	17.78	5.45	11.36	10.20	15.25	38.03	1.25	3.77	22.93
38	8.33	17.50	17.78	16.67	47.95	10.34	16.67	22.73	20.84
Average	**18.49**	**24.34**	**20.86**	**21.13**	**20.25**	**19.64**	**21.75**	**21.91**	**23.33**

Table 35. Number of data points in each scenario which fell into cluster 3

No.	Sc. 1	Sc. 2	Sc. 3	Sc. 4	Sc. 5	Sc. 6	Sc. 7	Sc. 8	Baseline
1	22	2	12	18	3	16	21	9	247
2	3	3	4	1	1	0	9	63	88
3	40	38	26	24	37	13	23	55	323
4	0	28	31	54	24	34	42	84	207
5	7	2	2	1	0	2	8	0	182
6	10	5	12	23	16	6	14	9	347
7	24	26	31	71	33	20	36	38	311
8	2	15	4	30	37	4	8	1	50
9	21	10	8	15	10	12	28	18	218
10	11	2	12	0	0	1	1	6	72
11	21	17	23	49	40	47	43	62	271
12	0	3	21	6	16	3	5	2	204
13	9	3	25	0	31	33	0	29	257
14	10	13	19	35	47	42	23	7	272
15	18	3	6	4	37	16	70	3	188
16	41	0	35	11	13	0	59	0	237
17	39	31	40	18	17	23	19	48	351
18	13	1	13	25	16	7	11	28	245
19	0	0	1	0	1	1	8	0	317
20	6	18	26	19	62	52	21	27	222

21	5	6	11	25	13	9	16	5	122
22	16	15	7	15	25	10	18	11	221
23	21	24	24	6	20	18	28	48	210
24	1	2	7	1	9	7	3	5	80
25	24	12	6	14	25	20	20	37	275
26	14	16	18	34	48	17	15	22	288
27	17	4	6	4	4	9	5	13	160
28	34	23	22	0	1	31	14	79	407
29	42	20	5	19	37	71	45	7	268
30	17	27	14	59	38	44	19	30	213
31	5	15	12	18	10	18	34	37	170
32	8	9	1	15	3	8	5	7	168
33	1	29	4	88	79	1	1	1	324
34	10	1	6	10	12	1	11	7	124
35	12	14	5	5	7	17	23	6	106
36	0	33	32	38	37	43	34	5	222
37	7	27	6	8	11	17	20	3	268
38	6	3	1	7	1	3	3	3	160

Table 36. Proportion of data in each scenario which fell into cluster 3

No.	Sc. 1	Sc. 2	Sc. 3	Sc. 4	Sc. 5	Sc. 6	Sc. 7	Sc. 8	Baseline
1	46.81	1.98	12.00	17.82	3.00	15.84	21.00	8.82	30.68
2	5.45	3.00	4.04	1.00	1.00	0.00	9.09	61.17	11.00
3	68.97	37.62	25.74	24.00	36.63	12.87	23.00	54.46	40.07
4	0.00	27.72	31.00	53.47	23.76	33.66	41.58	74.34	25.27
5	17.50	2.00	1.92	1.00	0.00	2.00	7.92	0.00	22.58
6	20.83	4.95	12.00	23.00	16.00	6.06	14.00	9.09	43.48
7	57.14	26.00	31.00	70.30	33.00	19.80	36.00	37.62	38.63
8	4.44	15.00	3.96	29.70	35.92	3.96	7.92	0.99	6.18
9	38.18	9.71	7.77	14.85	9.90	11.76	27.72	17.65	26.78
10	31.43	1.96	11.88	0.00	0.00	0.99	0.97	5.83	8.85
11	42.00	17.00	23.47	47.12	40.40	48.45	43.00	62.00	33.66
12	0.00	2.97	19.81	6.00	16.00	3.00	5.00	1.98	25.25
13	21.43	3.00	25.00	0.00	31.00	33.00	0.00	29.00	32.13
14	24.39	13.13	19.19	27.34	45.19	36.21	21.90	6.93	31.89
15	23.68	3.00	6.00	3.96	35.92	15.84	70.00	3.00	23.33
16	80.39	0.00	35.00	10.89	13.00	0.00	59.00	0.00	29.51
17	73.58	31.00	40.40	18.00	17.00	22.33	19.00	47.06	43.66
18	22.41	1.00	12.62	25.25	15.09	6.93	11.00	27.72	30.17
19	0.00	0.00	1.00	0.00	1.00	1.00	8.00	0.00	39.63
20	13.04	18.00	26.26	19.00	63.27	53.61	20.59	27.55	27.75

21	11.36	5.94	11.00	25.00	12.87	8.91	15.84	4.81	15.06
22	39.02	15.15	7.00	15.00	24.75	10.10	18.00	11.00	27.63
23	48.84	24.00	24.00	5.88	19.23	18.00	27.72	47.52	25.96
24	2.56	1.94	7.07	1.00	9.00	7.00	3.00	4.95	9.95
25	31.17	12.00	5.83	14.00	25.00	19.80	19.80	33.64	33.70
26	28.57	16.16	18.00	34.00	48.00	17.00	15.15	22.00	35.96
27	41.46	4.00	5.94	4.00	4.00	8.82	5.00	13.00	19.90
28	66.67	21.10	22.00	0.00	0.98	31.00	14.00	63.20	48.39
29	80.77	20.20	5.00	19.00	37.00	71.00	38.46	7.00	32.76
30	23.61	26.21	14.00	59.00	38.00	44.00	19.00	30.30	26.56
31	10.87	15.00	11.88	18.00	10.10	18.00	34.00	36.63	21.20
32	13.33	9.09	0.98	15.00	2.97	8.00	5.00	6.93	20.90
33	2.08	29.00	3.96	88.00	78.22	1.00	1.00	0.99	40.30
34	12.66	1.01	6.19	9.90	10.08	1.00	11.11	7.14	15.14
35	27.27	13.73	4.95	5.00	7.00	16.83	23.00	5.94	13.17
36	0.00	33.33	32.00	38.00	36.27	43.00	34.00	5.00	27.34
37	15.56	26.21	5.94	7.92	10.58	16.83	20.00	3.00	33.05
38	12.50	3.00	1.00	6.73	0.99	3.00	2.97	3.00	19.85
Average	**27.89**	**13.06**	**14.13**	**19.95**	**21.37**	**17.65**	**19.84**	**20.56**	**27.30**

Table 37. Number of data points in each scenario which fell into cluster 4

No.	Sc. 1	Sc. 2	Sc. 3	Sc. 4	Sc. 5	Sc. 6	Sc. 7	Sc. 8	Baseline
1	6	10	7	6	7	1	2	4	132
2	33	3	3	38	46	14	25	1	87
3	8	0	4	12	10	19	0	6	193
4	59	6	35	12	41	29	8	18	216
5	18	16	20	28	21	13	18	31	203
6	18	8	12	24	12	13	4	15	322
7	3	3	7	1	10	6	6	4	205
8	20	8	13	3	11	40	4	7	239
9	4	1	2	0	0	35	10	1	189
10	14	9	18	34	17	7	23	29	250
11	7	7	8	10	7	7	13	22	56
12	1	3	0	7	6	1	6	1	261
13	19	22	3	25	0	4	54	1	203
14	2	4	7	13	9	6	8	12	90
15	47	47	52	90	42	48	19	8	95
16	6	1	6	84	35	2	14	1	174
17	9	6	26	10	18	21	11	26	133
18	25	19	9	20	20	16	22	16	174
19	15	24	48	17	97	77	88	12	25
20	11	0	6	16	6	7	15	30	253

21	22	11	25	20	23	52	37	40	255
22	14	16	13	13	19	19	18	10	224
23	4	7	4	33	15	26	29	28	292
24	18	11	14	41	25	15	22	25	328
25	5	10	13	10	8	7	12	12	107
26	28	13	17	15	0	29	21	20	277
27	6	7	4	12	13	26	17	13	230
28	2	1	3	8	6	2	5	16	64
29	2	2	12	0	1	1	0	18	171
30	7	4	4	22	0	3	10	5	195
31	5	2	0	1	4	1	12	1	119
32	10	8	19	67	14	7	25	31	207
33	22	0	0	0	0	29	33	67	155
34	14	5	10	5	43	80	28	69	43
35	8	9	3	32	20	46	7	46	163
36	0	10	13	18	39	26	19	7	264
37	5	14	8	8	5	3	19	13	129
38	7	7	4	4	7	7	9	6	331

Table 38. Proportion of data in each scenario which fell into cluster 4

No.	Sc. 1	Sc. 2	Sc. 3	Sc. 4	Sc. 5	Sc. 6	Sc. 7	Sc. 8	Baseline
1	12.77	30.30	19.44	12.24	15.56	2.04	3.13	7.69	16.40
2	60.00	7.69	7.14	55.07	83.64	24.14	43.10	1.56	10.88
3	13.79	0.00	8.70	20.69	18.18	21.11	0.00	8.57	23.95
4	84.29	10.17	34.31	12.90	46.59	31.52	9.20	15.79	26.37
5	45.00	41.03	50.00	58.33	36.21	27.66	28.13	52.54	25.19
6	37.50	14.55	16.90	34.29	16.90	19.12	3.88	23.44	40.35
7	7.14	7.69	14.58	0.96	16.95	9.09	9.09	6.35	25.47
8	44.44	17.39	26.00	3.70	14.29	50.63	6.35	10.45	29.54
9	7.27	2.38	4.65	0.00	0.00	44.30	13.89	1.59	23.22
10	40.00	24.32	33.96	68.00	34.69	12.73	46.94	60.42	30.71
11	14.00	14.58	16.00	11.90	10.61	8.33	15.48	21.36	6.96
12	2.63	7.89	0.00	10.45	11.32	1.96	11.76	1.89	32.30
13	45.24	57.89	7.69	42.37	0.00	8.00	83.08	1.54	25.38
14	4.88	9.76	17.07	15.66	13.04	9.23	14.04	21.05	10.55
15	61.84	87.04	86.67	93.75	47.19	72.73	21.11	11.27	11.79
16	11.76	1.89	10.34	80.77	54.69	1.92	13.46	0.97	21.67
17	16.98	12.24	36.62	9.80	22.50	29.17	22.45	29.55	16.54
18	43.10	41.30	20.45	31.75	35.09	24.24	36.67	26.67	21.43
19	19.74	39.34	75.00	19.54	89.81	78.57	87.13	24.49	3.13
20	23.91	0.00	13.64	22.86	8.00	9.46	20.55	33.33	31.63

21	50.00	26.83	59.52	19.61	41.82	68.42	57.81	75.47	31.48
22	34.15	40.00	35.14	29.55	30.65	35.19	25.00	24.39	28.00
23	9.30	16.67	9.52	55.93	26.32	30.23	37.66	33.73	36.09
24	46.15	29.73	35.00	74.55	31.65	27.78	39.29	47.17	40.80
25	6.49	26.32	32.50	19.23	13.11	12.96	23.08	12.90	13.11
26	57.14	36.11	45.95	28.85	0.00	51.79	42.00	38.46	34.58
27	14.63	16.67	9.30	19.67	25.00	34.21	24.29	25.00	28.61
28	3.92	2.70	7.14	12.31	11.32	3.45	9.09	12.80	7.61
29	3.85	3.85	26.67	0.00	1.79	1.10	0.00	29.51	20.90
30	9.72	8.70	7.14	21.57	0.00	4.11	14.71	6.67	24.31
31	10.87	4.55	0.00	1.45	5.97	1.82	10.53	1.72	14.84
32	16.67	19.05	37.25	71.28	23.73	6.86	38.46	31.31	25.75
33	45.83	0.00	0.00	0.00	0.00	43.28	37.50	78.82	19.28
34	17.72	6.67	19.61	7.58	66.15	97.56	42.42	81.18	5.25
35	18.18	19.57	6.52	47.06	37.04	45.10	11.29	45.10	20.25
36	0.00	20.41	27.08	26.09	39.00	34.21	27.54	6.86	32.51
37	11.11	25.45	18.18	16.33	8.47	4.23	23.75	24.53	15.91
38	14.58	17.50	8.89	5.56	9.59	12.07	15.00	9.09	41.07
Avg	**25.44**	**19.69**	**23.28**	**27.94**	**24.92**	**26.32**	**25.50**	**24.87**	**22.99**

Case 2: Cluster all scenario data (excluding baseline data) of each participant

Table 39. Number of data points in each scenario which fell into cluster 1

No.	Sc. 1	Sc. 2	Sc. 3	Sc. 4	Sc. 5	Sc. 6	Sc. 7	Sc. 8
1	7	16	7	14	13	4	18	13
2	24	2	1	19	47	4	19	1
3	37	37	25	22	30	5	21	42
4	0	22	24	22	5	48	51	28
5	11	13	8	14	28	19	19	14
6	12	10	15	19	16	11	46	17
7	9	16	7	67	7	7	17	6
8	2	17	20	30	20	13	41	44
9	15	15	10	15	20	8	8	20
10	6	11	6	12	17	7	9	4
11	17	13	15	21	26	21	26	55
12	10	13	13	7	8	14	11	17
13	27	29	2	8	0	1	60	1
14	13	1	1	16	3	5	3	11
15	0	2	0	0	0	1	0	67
16	7	4	50	10	17	1	98	1
17	8	5	4	82	30	21	13	2
18	12	16	6	10	15	28	20	10

19	47	35	9	66	0	6	0	12
20	1	8	20	12	49	25	8	13
21	10	4	17	7	18	19	25	13
22	15	14	6	12	20	6	13	9
23	12	8	7	33	15	30	26	28
24	14	13	21	1	37	21	22	17
25	5	10	16	13	8	8	12	15
26	12	15	19	34	42	16	15	20
27	8	3	4	5	3	4	2	4
28	14	12	13	2	10	22	4	22
29	1	17	7	84	16	1	13	3
30	12	23	10	65	30	24	8	21
31	7	5	0	34	17	8	52	3
32	6	10	20	12	23	9	16	50
33	27	0	0	0	0	38	34	75
34	4	5	5	3	33	76	1	51
35	16	22	4	21	24	19	29	42
36	0	5	13	28	33	22	30	81
37	27	3	17	30	31	12	12	36
38	26	21	17	42	5	26	27	34

Table 40. Proportion of data in each scenario which fell into cluster 1

No.	Sc. 1	Sc. 2	Sc. 3	Sc. 4	Sc. 5	Sc. 6	Sc. 7	Sc. 8
1	14.89	48.48	19.44	28.57	28.89	8.16	28.13	25.00
2	43.64	5.13	2.38	27.54	85.45	6.90	32.76	1.56
3	63.79	82.22	54.35	37.93	54.55	5.56	37.50	60.00
4	0.00	37.29	23.53	23.66	5.68	52.17	58.62	24.56
5	27.50	33.33	20.00	29.17	48.28	40.43	29.69	23.73
6	25.00	18.18	21.13	27.14	22.54	16.18	44.66	26.56
7	21.43	41.03	14.58	64.42	11.86	10.61	25.76	9.52
8	4.44	36.96	40.00	37.04	25.97	16.46	65.08	65.67
9	27.27	35.71	23.26	23.08	27.40	10.13	11.11	31.75
10	17.14	29.73	11.32	24.00	34.69	12.73	18.37	8.33
11	34.00	27.08	30.00	25.00	39.39	25.00	30.95	53.40
12	26.32	34.21	30.95	10.45	15.09	27.45	21.57	32.08
13	64.29	76.32	5.13	13.56	0.00	2.00	92.31	1.54
14	31.71	2.44	2.44	19.28	4.35	7.69	5.26	19.30
15	0.00	3.70	0.00	0.00	0.00	1.52	0.00	94.37
16	13.73	7.55	86.21	9.62	26.56	0.96	94.23	0.97
17	15.09	10.20	5.63	80.39	37.50	29.17	26.53	2.27
18	20.69	34.78	13.64	15.87	26.32	42.42	33.33	16.67
19	61.84	57.38	14.06	75.86	0.00	6.12	0.00	24.49
20	2.17	18.18	45.45	17.14	65.33	33.78	10.96	14.44

21	22.73	9.76	40.48	6.86	32.73	25.00	39.06	24.53
22	36.59	35.00	16.22	27.27	32.26	11.11	18.06	21.95
23	27.91	19.05	16.67	55.93	26.32	34.88	33.77	33.73
24	35.90	35.14	52.50	1.82	46.84	38.89	39.29	32.08
25	6.49	26.32	40.00	25.00	13.11	14.81	23.08	16.13
26	24.49	41.67	51.35	65.38	85.71	28.57	30.00	38.46
27	19.51	7.14	9.30	8.20	5.77	5.26	2.86	7.69
28	27.45	32.43	30.95	3.08	18.87	37.93	7.27	17.60
29	1.92	32.69	15.56	82.35	28.57	1.10	21.67	4.92
30	16.67	50.00	17.86	63.73	52.63	32.88	11.76	28.00
31	15.22	11.36	0.00	49.28	25.37	14.55	45.61	5.17
32	10.00	23.81	39.22	12.77	38.98	8.82	24.62	50.51
33	56.25	0.00	0.00	0.00	0.00	56.72	38.64	88.24
34	5.06	6.67	9.80	4.55	50.77	92.68	1.52	60.00
35	36.36	47.83	8.70	30.88	44.44	18.63	46.77	41.18
36	0.00	10.20	27.08	40.58	33.00	28.95	43.48	79.41
37	60.00	5.45	38.64	61.22	52.54	16.90	15.00	67.92
38	54.17	52.50	37.78	58.33	6.85	44.83	45.00	51.52
Avg	**25.57**	**28.60**	**24.09**	**31.24**	**30.38**	**22.84**	**30.38**	**31.72**

Table 41. Number of data points in each scenario which fell into cluster 2

No.	Sc. 1	Sc. 2	Sc. 3	Sc. 4	Sc. 5	Sc. 6	Sc. 7	Sc. 8
1	16	4	8	1	16	19	26	20
2	7	2	9	3	3	0	15	62
3	8	2	2	3	10	38	15	16
4	1	28	29	52	15	29	36	80
5	4	7	12	4	9	9	8	15
6	6	21	28	8	25	26	42	13
7	16	10	12	0	21	40	22	18
8	38	1	6	15	10	24	7	11
9	24	7	12	10	12	11	31	18
10	11	9	22	27	11	6	17	34
11	9	10	14	33	21	38	22	15
12	1	4	14	8	27	1	7	5
13	0	0	0	49	0	0	3	0
14	18	22	14	23	13	13	22	27
15	1	10	50	3	32	41	8	2
16	1	49	3	16	19	103	0	0
17	9	6	24	11	17	22	12	24
18	26	0	22	21	16	15	17	16
19	9	12	18	8	8	65	75	18
20	7	8	11	9	19	33	12	17

21	15	10	6	57	9	14	6	6
22	5	15	9	15	14	20	30	19
23	18	24	13	20	33	40	5	4
24	4	7	2	5	2	9	9	13
25	26	14	6	15	25	19	22	37
26	1	13	13	16	3	29	14	15
27	4	6	9	15	15	9	28	15
28	35	21	22	0	1	33	10	81
29	30	16	2	3	13	75	15	1
30	41	4	13	21	4	7	11	12
31	25	15	21	13	36	16	15	6
32	39	27	27	0	30	83	9	35
33	16	10	59	9	4	28	52	8
34	47	65	25	32	0	1	8	3
35	12	3	2	9	8	20	4	12
36	0	9	8	14	28	18	15	15
37	8	5	6	5	10	27	4	3
38	5	4	4	3	3	1	3	6

Table 42. Proportion of data in each scenario which fell into cluster 2

No.	Sc. 1	Sc. 2	Sc. 3	Sc. 4	Sc. 5	Sc. 6	Sc. 7	Sc. 8
1	34.04	12.12	22.22	2.04	35.56	38.78	40.63	38.46
2	12.73	5.13	21.43	4.35	5.45	0.00	25.86	96.88
3	13.79	4.44	4.35	5.17	18.18	42.22	26.79	22.86
4	1.43	47.46	28.43	55.91	17.05	31.52	41.38	70.18
5	10.00	17.95	30.00	8.33	15.52	19.15	12.50	25.42
6	12.50	38.18	39.44	11.43	35.21	38.24	40.78	20.31
7	38.10	25.64	25.00	0.00	35.59	60.61	33.33	28.57
8	84.44	2.17	12.00	18.52	12.99	30.38	11.11	16.42
9	43.64	16.67	27.91	15.38	16.44	13.92	43.06	28.57
10	31.43	24.32	41.51	54.00	22.45	10.91	34.69	70.83
11	18.00	20.83	28.00	39.29	31.82	45.24	26.19	14.56
12	2.63	10.53	33.33	11.94	50.94	1.96	13.73	9.43
13	0.00	0.00	0.00	83.05	0.00	0.00	4.62	0.00
14	43.90	53.66	34.15	27.71	18.84	20.00	38.60	47.37
15	1.32	18.52	83.33	3.13	35.96	62.12	8.89	2.82
16	1.96	92.45	5.17	15.38	29.69	99.04	0.00	0.00
17	16.98	12.24	33.80	10.78	21.25	30.56	24.49	27.27
18	44.83	0.00	50.00	33.33	28.07	22.73	28.33	26.67
19	11.84	19.67	28.13	9.20	7.41	66.33	74.26	36.73
20	15.22	18.18	25.00	12.86	25.33	44.59	16.44	18.89

21	34.09	24.39	14.29	55.88	16.36	18.42	9.38	11.32
22	12.20	37.50	24.32	34.09	22.58	37.04	41.67	46.34
23	41.86	57.14	30.95	33.90	57.89	46.51	6.49	4.82
24	10.26	18.92	5.00	9.09	2.53	16.67	16.07	24.53
25	33.77	36.84	15.00	28.85	40.98	35.19	42.31	39.78
26	2.04	36.11	35.14	30.77	6.12	51.79	28.00	28.85
27	9.76	14.29	20.93	24.59	28.85	11.84	40.00	28.85
28	68.63	56.76	52.38	0.00	1.89	56.90	18.18	64.80
29	57.69	30.77	4.44	2.94	23.21	82.42	25.00	1.64
30	56.94	8.70	23.21	20.59	7.02	9.59	16.18	16.00
31	54.35	34.09	42.00	18.84	53.73	29.09	13.16	10.34
32	65.00	64.29	52.94	0.00	50.85	81.37	13.85	35.35
33	33.33	21.28	95.16	8.82	4.60	41.79	59.09	9.41
34	59.49	86.67	49.02	48.48	0.00	1.22	12.12	3.53
35	27.27	6.52	4.35	13.24	14.81	19.61	6.45	11.76
36	0.00	18.37	16.67	20.29	28.00	23.68	21.74	14.71
37	17.78	9.09	13.64	10.20	16.95	38.03	5.00	5.66
38	10.42	10.00	8.89	4.17	4.11	1.72	5.00	9.09
Avg	**27.20**	**26.63**	**28.46**	**20.70**	**22.22**	**33.71**	**24.35**	**25.50**

Table 43. Number of data points in each scenario which fell into cluster 3

No.	Sc. 1	Sc. 2	Sc. 3	Sc. 4	Sc. 5	Sc. 6	Sc. 7	Sc. 8
1	9	11	17	30	14	14	10	14
2	8	23	25	10	2	32	10	1
3	0	2	2	10	3	6	5	1
4	1	8	33	10	12	13	0	2
5	4	13	14	28	17	10	8	25
6	7	0	9	7	8	9	7	9
7	4	2	4	26	9	6	9	11
8	0	13	2	29	31	0	8	1
9	6	1	3	4	2	50	20	2
10	5	15	21	5	10	39	20	10
11	5	7	8	10	8	3	13	20
12	11	10	10	13	13	20	11	18
13	8	7	1	1	4	6	1	32
14	1	5	6	12	8	5	10	11
15	30	13	7	7	56	17	77	1
16	41	0	4	78	28	0	6	1
17	1	8	6	3	7	10	11	13
18	12	0	12	20	17	5	11	28
19	8	1	1	2	4	1	13	12
20	23	22	7	25	5	11	33	28

21	5	23	5	34	18	6	6	5
22	19	6	20	13	18	17	21	10
23	7	5	15	1	5	2	33	48
24	5	4	2	8	21	2	6	7
25	25	8	16	20	14	24	12	21
26	2	3	2	0	1	10	13	5
27	21	23	26	18	14	31	19	14
28	1	2	5	56	36	0	36	7
29	4	9	34	9	4	1	1	51
30	8	4	8	1	4	8	5	4
31	10	21	25	13	12	26	35	42
32	9	0	3	70	3	5	36	5
33	1	2	1	11	70	0	0	0
34	10	2	6	12	15	1	12	9
35	10	21	2	18	19	38	28	40
36	54	0	1	0	0	0	0	0
37	8	0	17	3	13	27	61	7
38	14	9	17	14	10	29	22	11

Table 44. Proportion of data in each scenario which fell into cluster 3

No.	Sc. 1	Sc. 2	Sc. 3	Sc. 4	Sc. 5	Sc. 6	Sc. 7	Sc. 8
1	19.15	33.33	47.22	61.22	31.11	28.57	15.63	26.92
2	14.55	58.97	59.52	14.49	3.64	55.17	17.24	1.56
3	0.00	4.44	4.35	17.24	5.45	6.67	8.93	1.43
4	1.43	13.56	32.35	10.75	13.64	14.13	0.00	1.75
5	10.00	33.33	35.00	58.33	29.31	21.28	12.50	42.37
6	14.58	0.00	12.68	10.00	11.27	13.24	6.80	14.06
7	9.52	5.13	8.33	25.00	15.25	9.09	13.64	17.46
8	0.00	28.26	4.00	35.80	40.26	0.00	12.70	1.49
9	10.91	2.38	6.98	6.15	2.74	63.29	27.78	3.17
10	14.29	40.54	39.62	10.00	20.41	70.91	40.82	20.83
11	10.00	14.58	16.00	11.90	12.12	3.57	15.48	19.42
12	28.95	26.32	23.81	19.40	24.53	39.22	21.57	33.96
13	19.05	18.42	2.56	1.69	7.41	12.00	1.54	49.23
14	2.44	12.20	14.63	14.46	11.59	7.69	17.54	19.30
15	39.47	24.07	11.67	7.29	62.92	25.76	85.56	1.41
16	80.39	0.00	6.90	75.00	43.75	0.00	5.77	0.97
17	1.89	16.33	8.45	2.94	8.75	13.89	22.45	14.77
18	20.69	0.00	27.27	31.75	29.82	7.58	18.33	46.67
19	10.53	1.64	1.56	2.30	3.70	1.02	12.87	24.49
20	50.00	50.00	15.91	35.71	6.67	14.86	45.21	31.11

21	11.36	56.10	11.90	33.33	32.73	7.89	9.38	9.43
22	46.34	15.00	54.05	29.55	29.03	31.48	29.17	24.39
23	16.28	11.90	35.71	1.69	8.77	2.33	42.86	57.83
24	12.82	10.81	5.00	14.55	26.58	3.70	10.71	13.21
25	32.47	21.05	40.00	38.46	22.95	44.44	23.08	22.58
26	4.08	8.33	5.41	0.00	2.04	17.86	26.00	9.62
27	51.22	54.76	60.47	29.51	26.92	40.79	27.14	26.92
28	1.96	5.41	11.90	86.15	67.92	0.00	65.45	5.60
29	7.69	17.31	75.56	8.82	7.14	1.10	1.67	83.61
30	11.11	8.70	14.29	0.98	7.02	10.96	7.35	5.33
31	21.74	47.73	50.00	18.84	17.91	47.27	30.70	72.41
32	15.00	0.00	5.88	74.47	5.08	4.90	55.38	5.05
33	2.08	4.26	1.61	10.78	80.46	0.00	0.00	0.00
34	12.66	2.67	11.76	18.18	23.08	1.22	18.18	10.59
35	22.73	45.65	4.35	26.47	35.19	37.25	45.16	39.22
36	100.00	0.00	2.08	0.00	0.00	0.00	0.00	0.00
37	17.78	0.00	38.64	6.12	22.03	38.03	76.25	13.21
38	29.17	22.50	37.78	19.44	13.70	50.00	36.67	16.67
Avg	**20.38**	**18.83**	**22.24**	**22.86**	**21.39**	**19.66**	**23.88**	**20.74**

Table 45. Number of data points in each scenario which fell into cluster 4

No.	Sc. 1	Sc. 2	Sc. 3	Sc. 4	Sc. 5	Sc. 6	Sc. 7	Sc. 8
1	15	2	4	4	2	12	10	5
2	16	12	7	37	3	22	14	0
3	13	4	17	23	12	41	15	11
4	68	1	16	9	56	2	0	4
5	21	6	6	2	4	9	29	5
6	23	24	19	36	22	22	8	25
7	13	11	25	11	22	13	18	28
8	5	15	22	7	16	42	7	11
9	10	19	18	36	39	10	13	23
10	13	2	4	6	11	3	3	0
11	19	18	13	20	11	22	23	13
12	16	11	5	39	5	16	22	13
13	7	2	36	1	50	43	1	32
14	9	13	20	32	45	42	22	8
15	45	29	3	86	1	7	5	1
16	2	0	1	0	0	0	0	101
17	35	30	37	6	26	19	13	49
18	8	30	4	12	9	18	12	6
19	12	13	36	11	96	26	13	7
20	15	6	6	24	2	5	20	32

21	14	4	14	4	10	37	27	29
22	2	5	2	4	10	11	8	3
23	6	5	7	5	4	14	13	3
24	16	13	15	41	19	22	19	16
25	21	6	2	4	14	3	6	20
26	34	5	3	2	3	1	8	12
27	8	10	4	23	20	32	21	19
28	1	2	2	7	6	3	5	15
29	17	10	2	6	23	14	31	6
30	11	15	25	15	19	34	44	38
31	4	3	4	9	2	5	12	7
32	6	5	1	12	3	5	4	9
33	4	35	2	82	13	1	2	2
34	18	3	15	19	17	4	45	22
35	6	0	38	20	3	25	1	8
36	0	35	26	27	39	36	24	6
37	2	47	4	11	5	5	3	7
38	3	6	7	13	55	2	8	15

Table 46. Proportion of data in each scenario which fell into cluster 3

No.	Sc. 1	Sc. 2	Sc. 3	Sc. 4	Sc. 5	Sc. 6	Sc. 7	Sc. 8
1	31.91	6.06	11.11	8.16	4.44	24.49	15.63	9.62
2	29.09	30.77	16.67	53.62	5.45	37.93	24.14	0.00
3	22.41	8.89	36.96	39.66	21.82	45.56	26.79	15.71
4	97.14	1.69	15.69	9.68	63.64	2.17	0.00	3.51
5	52.50	15.38	15.00	4.17	6.90	19.15	45.31	8.47
6	47.92	43.64	26.76	51.43	30.99	32.35	7.77	39.06
7	30.95	28.21	52.08	10.58	37.29	19.70	27.27	44.44
8	11.11	32.61	44.00	8.64	20.78	53.16	11.11	16.42
9	18.18	45.24	41.86	55.38	53.42	12.66	18.06	36.51
10	37.14	5.41	7.55	12.00	22.45	5.45	6.12	0.00
11	38.00	37.50	26.00	23.81	16.67	26.19	27.38	12.62
12	42.11	28.95	11.90	58.21	9.43	31.37	43.14	24.53
13	16.67	5.26	92.31	1.69	92.59	86.00	1.54	49.23
14	21.95	31.71	48.78	38.55	65.22	64.62	38.60	14.04
15	59.21	53.70	5.00	89.58	1.12	10.61	5.56	1.41
16	3.92	0.00	1.72	0.00	0.00	0.00	0.00	98.06
17	66.04	61.22	52.11	5.88	32.50	26.39	26.53	55.68
18	13.79	65.22	9.09	19.05	15.79	27.27	20.00	10.00
19	15.79	21.31	56.25	12.64	88.89	26.53	12.87	14.29
20	32.61	13.64	13.64	34.29	2.67	6.76	27.40	35.56

21	31.82	9.76	33.33	3.92	18.18	48.68	42.19	54.72
22	4.88	12.50	5.41	9.09	16.13	20.37	11.11	7.32
23	13.95	11.90	16.67	8.47	7.02	16.28	16.88	3.61
24	41.03	35.14	37.50	74.55	24.05	40.74	33.93	30.19
25	27.27	15.79	5.00	7.69	22.95	5.56	11.54	21.51
26	69.39	13.89	8.11	3.85	6.12	1.79	16.00	23.08
27	19.51	23.81	9.30	37.70	38.46	42.11	30.00	36.54
28	1.96	5.41	4.76	10.77	11.32	5.17	9.09	12.00
29	32.69	19.23	4.44	5.88	41.07	15.38	51.67	9.84
30	15.28	32.61	44.64	14.71	33.33	46.58	64.71	50.67
31	8.70	6.82	8.00	13.04	2.99	9.09	10.53	12.07
32	10.00	11.90	1.96	12.77	5.08	4.90	6.15	9.09
33	8.33	74.47	3.23	80.39	14.94	1.49	2.27	2.35
34	22.78	4.00	29.41	28.79	26.15	4.88	68.18	25.88
35	13.64	0.00	82.61	29.41	5.56	24.51	1.61	7.84
36	0.00	71.43	54.17	39.13	39.00	47.37	34.78	5.88
37	4.44	85.45	9.09	22.45	8.47	7.04	3.75	13.21
38	6.25	15.00	15.56	18.06	75.34	3.45	13.33	22.73
Avg	**26.85**	**25.93**	**25.20**	**25.20**	**26.01**	**23.78**	**21.39**	**22.04**

Gaussian model

Case 1: Cluster all data (scenario and baseline combined) of each participant

Table 47. Number of data points in each scenario which were detected as anomaly

No.	Sc. 1	Sc. 2	Sc. 3	Sc. 4	Sc. 5	Sc. 6	Sc. 7	Sc. 8	Baseline
1	7	6	14	5	8	7	14	10	188
2	17	8	9	17	16	15	16	19	141
3	5	0	5	7	10	10	3	13	152
4	12	14	16	31	27	48	10	28	183
5	4	3	7	3	4	7	2	7	165
6	21	5	21	18	17	18	32	16	138
7	7	1	4	11	16	16	8	11	216
8	9	9	7	16	19	14	9	4	138
9	14	2	9	8	9	8	12	8	196
10	12	6	21	7	12	12	16	6	206
11	2	5	3	8	7	5	6	11	184
12	10	10	7	15	12	9	10	7	218
13	12	10	6	21	13	11	9	15	186
14	15	3	4	10	6	11	8	16	206
15	10	14	16	22	19	35	19	35	180
16	16	8	9	27	5	7	6	33	219
17	15	15	15	28	23	13	24	23	148

18	15	20	7	13	13	28	23	11	149
19	24	20	7	20	20	17	29	12	159
20	5	2	2	9	8	8	12	3	192
21	14	12	3	21	10	16	10	9	220
22	5	13	15	11	15	8	11	11	173
23	5	8	5	6	8	11	14	12	215
24	13	6	8	10	20	14	16	22	160
25	22	13	7	7	8	6	11	16	177
26	7	6	2	3	3	6	8	2	205
27	12	6	9	11	7	8	11	14	151
28	4	8	9	25	19	14	14	23	187
29	4	6	13	14	12	4	6	14	213
30	16	4	9	23	1	10	9	11	182
31	7	4	2	12	6	3	19	10	223
32	6	7	2	24	11	8	11	9	131
33	7	7	13	6	9	6	8	16	204
34	29	20	10	8	12	28	9	32	133
35	19	8	18	16	11	31	18	11	197
36	35	23	8	28	29	13	14	14	167
37	5	13	6	8	7	16	16	9	234
38	9	5	10	21	12	12	11	6	235

Table 48. Proportion of data in each scenario which were detected as anomaly points

No.	Sc. 1	Sc. 2	Sc. 3	Sc. 4	Sc. 5	Sc. 6	Sc. 7	Sc. 8	Baseline
1	14.89	18.18	38.89	10.20	17.78	14.29	21.88	19.23	23.35
2	30.91	20.51	21.43	24.64	29.09	25.86	27.59	29.69	17.63
3	8.62	0.00	10.87	12.07	18.18	11.11	5.36	18.57	18.86
4	17.14	23.73	15.69	33.33	30.68	52.17	11.49	24.56	22.34
5	10.00	7.69	17.50	6.25	6.90	14.89	3.13	11.86	20.47
6	43.75	9.09	29.58	25.71	23.94	26.47	31.07	25.00	17.29
7	16.67	2.56	8.33	10.58	27.12	24.24	12.12	17.46	26.83
8	20.00	19.57	14.00	19.75	24.68	17.72	14.29	5.97	17.06
9	25.45	4.76	20.93	12.31	12.33	10.13	16.67	12.70	24.08
10	34.29	16.22	39.62	14.00	24.49	21.82	32.65	12.50	25.31
11	4.00	10.42	6.00	9.52	10.61	5.95	7.14	10.68	22.86
12	26.32	26.32	16.67	22.39	22.64	17.65	19.61	13.21	26.98
13	28.57	26.32	15.38	35.59	24.07	22.00	13.85	23.08	23.25
14	36.59	7.32	9.76	12.05	8.70	16.92	14.04	28.07	24.15
15	13.16	25.93	26.67	22.92	21.35	53.03	21.11	49.30	22.33
16	31.37	15.09	15.52	25.96	7.81	6.73	5.77	32.04	27.27
17	28.30	30.61	21.13	27.45	28.75	18.06	48.98	26.14	18.41
18	25.86	43.48	15.91	20.63	22.81	42.42	38.33	18.33	18.35
19	31.58	32.79	10.94	22.99	18.52	17.35	28.71	24.49	19.88
20	10.87	4.55	4.55	12.86	10.67	10.81	16.44	3.33	24.00

21	31.82	29.27	7.14	20.59	18.18	21.05	15.63	16.98	27.16
22	12.20	32.50	40.54	25.00	24.19	14.81	15.28	26.83	21.63
23	11.63	19.05	11.90	10.17	14.04	12.79	18.18	14.46	26.58
24	33.33	16.22	20.00	18.18	25.32	25.93	28.57	41.51	19.90
25	28.57	34.21	17.50	13.46	13.11	11.11	21.15	17.20	21.69
26	14.29	16.67	5.41	5.77	6.12	10.71	16.00	3.85	25.59
27	29.27	14.29	20.93	18.03	13.46	10.53	15.71	26.92	18.78
28	7.84	21.62	21.43	38.46	35.85	24.14	25.45	18.40	22.24
29	7.69	11.54	28.89	13.73	21.43	4.40	10.00	22.95	26.04
30	22.22	8.70	16.07	22.55	1.75	13.70	13.24	14.67	22.69
31	15.22	9.09	4.00	17.39	8.96	5.45	16.67	17.24	27.81
32	10.00	16.67	3.92	25.53	18.64	7.84	16.92	9.09	16.29
33	14.58	14.89	20.97	5.88	10.34	8.96	9.09	18.82	25.37
34	36.71	26.67	19.61	12.12	18.46	34.15	13.64	37.65	16.24
35	43.18	17.39	39.13	23.53	20.37	30.39	29.03	10.78	24.47
36	64.81	46.94	16.67	40.58	29.00	17.11	20.29	13.73	20.57
37	11.11	23.64	13.64	16.33	11.86	22.54	20.00	16.98	28.85
38	18.75	12.50	22.22	29.17	16.44	20.69	18.33	9.09	29.16
Average	**22.94**	**18.87**	**18.14**	**19.41**	**18.39**	**19.10**	**18.77**	**19.56**	**22.68**

Case 2: Cluster all scenario data (excluding baseline data) of each participant

Table 49. Number of data points in each scenario which were detected as anomaly

No.	Sc. 1	Sc. 2	Sc. 3	Sc. 4	Sc. 5	Sc. 6	Sc. 7	Sc. 8
1	8	5	15	11	13	6	19	11
2	21	14	14	24	21	21	17	21
3	7	3	7	15	12	13	7	9
4	18	12	34	31	37	40	8	17
5	6	9	11	6	8	12	11	13
6	23	6	14	19	14	15	33	14
7	11	3	8	16	14	14	14	14
8	9	12	14	17	20	18	12	8
9	22	7	8	6	13	20	18	10
10	16	5	18	12	15	13	14	8
11	10	11	5	13	16	14	14	22
12	9	8	7	16	27	9	11	7
13	19	11	14	16	15	24	8	18
14	14	7	11	18	18	12	11	22
15	12	18	16	26	16	33	14	18
16	22	6	16	24	18	22	22	15
17	14	17	21	31	38	20	24	29
18	15	21	16	10	16	20	23	14

19	21	12	9	31	24	20	33	15
20	10	8	7	12	12	20	14	13
21	18	14	3	19	12	14	15	11
22	8	4	11	9	15	18	14	13
23	7	9	7	13	9	19	23	18
24	13	7	9	12	20	19	22	26
25	24	16	10	13	12	7	10	18
26	11	6	3	2	11	10	15	7
27	13	12	14	17	10	18	12	15
28	8	14	8	26	22	14	17	35
29	4	7	11	16	18	14	4	19
30	30	11	13	26	7	17	16	18
31	13	5	3	17	10	7	27	5
32	15	12	5	28	15	16	13	16
33	8	12	19	12	13	11	14	29
34	27	16	14	14	25	17	16	28
35	23	11	23	20	15	37	21	27
36	16	15	10	29	37	18	16	22
37	7	19	8	9	11	21	20	13
38	11	4	7	20	21	13	11	15

Table 50. Proportion of data in each scenario which were detected as anomaly points

No.	Sc. 1	Sc. 2	Sc. 3	Sc. 4	Sc. 5	Sc. 6	Sc. 7	Sc. 8
1	17.02	15.15	41.67	22.45	28.89	12.24	29.69	21.15
2	38.18	35.90	33.33	34.78	38.18	36.21	29.31	32.81
3	12.07	6.67	15.22	25.86	21.82	14.44	12.50	12.86
4	25.71	20.34	33.33	33.33	42.05	43.48	9.20	14.91
5	15.00	23.08	27.50	12.50	13.79	25.53	17.19	22.03
6	47.92	10.91	19.72	27.14	19.72	22.06	32.04	21.88
7	26.19	7.69	16.67	15.38	23.73	21.21	21.21	22.22
8	20.00	26.09	28.00	20.99	25.97	22.78	19.05	11.94
9	40.00	16.67	18.60	9.23	17.81	25.32	25.00	15.87
10	45.71	13.51	33.96	24.00	30.61	23.64	28.57	16.67
11	20.00	22.92	10.00	15.48	24.24	16.67	16.67	21.36
12	23.68	21.05	16.67	23.88	50.94	17.65	21.57	13.21
13	45.24	28.95	35.90	27.12	27.78	48.00	12.31	27.69
14	34.15	17.07	26.83	21.69	26.09	18.46	19.30	38.60
15	15.79	33.33	26.67	27.08	17.98	50.00	15.56	25.35
16	43.14	11.32	27.59	23.08	28.13	21.15	21.15	14.56
17	26.42	34.69	29.58	30.39	47.50	27.78	48.98	32.95
18	25.86	45.65	36.36	15.87	28.07	30.30	38.33	23.33
19	27.63	19.67	14.06	35.63	22.22	20.41	32.67	30.61
20	21.74	18.18	15.91	17.14	16.00	27.03	19.18	14.44

21	40.91	34.15	7.14	18.63	21.82	18.42	23.44	20.75
22	19.51	10.00	29.73	20.45	24.19	33.33	19.44	31.71
23	16.28	21.43	16.67	22.03	15.79	22.09	29.87	21.69
24	33.33	18.92	22.50	21.82	25.32	35.19	39.29	49.06
25	31.17	42.11	25.00	25.00	19.67	12.96	19.23	19.35
26	22.45	16.67	8.11	3.85	22.45	17.86	30.00	13.46
27	31.71	28.57	32.56	27.87	19.23	23.68	17.14	28.85
28	15.69	37.84	19.05	40.00	41.51	24.14	30.91	28.00
29	7.69	13.46	24.44	15.69	32.14	15.38	6.67	31.15
30	41.67	23.91	23.21	25.49	12.28	23.29	23.53	24.00
31	28.26	11.36	6.00	24.64	14.93	12.73	23.68	8.62
32	25.00	28.57	9.80	29.79	25.42	15.69	20.00	16.16
33	16.67	25.53	30.65	11.76	14.94	16.42	15.91	34.12
34	34.18	21.33	27.45	21.21	38.46	20.73	24.24	32.94
35	52.27	23.91	50.00	29.41	27.78	36.27	33.87	26.47
36	29.63	30.61	20.83	42.03	37.00	23.68	23.19	21.57
37	15.56	34.55	18.18	18.37	18.64	29.58	25.00	24.53
38	22.92	10.00	15.56	27.78	28.77	22.41	18.33	22.73
Average	**27.80**	**22.68**	**23.54**	**23.39**	**26.10**	**24.43**	**23.51**	**23.41**

3. Subjective ratings of stress

Table 51. Results for subjective ratings of stress during scenarios (the scale is from 1 for totally relaxed to 7 for too stressful)

No.	Sc. 1	Sc. 2	Sc. 3	Sc. 4	Sc. 5	Sc. 6	Sc. 7	Sc. 8
1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2
3	1	3	1	3	4	1	3	1
4	3	4	3	7	7	4	4	4
5	1	1	1	1	1	1	1	1
6	5	5	6	5	5	6	5	5
7	4	3	3	4	4	4	4	4
8	2	4	3	4	3	3	4	2
9	2	2	2	2	2	2	2	2
10	1	1	1	1	1	1	1	1
11	5	6	6	5	4	6	4	4
12	1	1	1	2	1	2	1	1
13	1	1	1	1	1	1	1	1
14	2	3	1	3	1	2	2	3
15	4	4	4	4	4	4	4	4
16	4	4	4	3	3	4	4	5
17	1	1	1	1	1	1	1	1
18	1	2	1	2	2	1	2	1
19	3	3	3	3	3	3	3	3

20	1	1	1	1	1	1	1	1
21	1	1	1	1	1	1	1	1
22	1	1	1	1	1	1	1	1
23	2	2	2	2	2	2	2	2
24	1	1	1	1	1	1	1	1
25	2	1	1	2	1	1	1	2
26	1	1	1	1	1	1	1	1
27	5	4	4	6	4	4	5	5
28	5	4	3	3	3	4	3	4
29	1	1	1	1	1	1	1	1
30	2	2	2	2	2	2	2	2
31	2	2	2	2	2	2	2	2
32	4	4	4	4	4	4	4	4
33	3	5	4	4	4	5	4	4
34	1	1	1	1	1	1	1	1
35	4	5	4	4	4	4	5	2
36	4	4	4	4	4	4	4	4
37	1	1	1	1	1	1	1	1
38	4	4	4	4	5	5	4	4