January 2015

DETECTION OF OPERATOR PERFORMANCE BREAKDOWN IN A MULTITASK ENVIRONMENT

Hyo-Sang Yoo
Purdue University

Graduate School Form 30
Updated 1/15/2015

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Hyo-Sang Yoo

Entitled
DETECTION OF OPERATOR PERFORMANCE BREAKDOWN IN A MULTITASK ENVIRONMENT

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Steven J. Landry
Chair

Paul U. Lee

Barrett Caldwell

Ji Soo Yi

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Steven J. Landry

Approved by: Abhijit Deshmukh 11/2/2015

Head of the Departmental Graduate Program Date

DETECTION OF OPERATOR PERFORMANCE BREAKDOWN IN A MULTITASK ENVIRONMENT

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Hyo-Sang Yoo

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2015

Purdue University

West Lafayette, Indiana

I would like to dedicate my dissertation work to my beloved family members who have shaped my life: My father (Seungha Yoo), my mother (Junghee Lee), and my sister (Jihae Yoo).

ACKNOWLEDGEMENTS

First and foremost I want to thank my advisor, Dr. Steven J. Landry for his excellent guidance, caring, and patience. He has been supportive since the days I began working as a Summer Undergraduate Research Fellowship (SURF) student in 2008. He has taught me how good science is done. Dr. Landry has supported me not only by providing an assistantship for many years, but helped me academically and emotionally as I was going through this hard journey. I am thankful for him being a good role model for an engineer, mentor, and teacher. I will thank him forever for the opportunities he gave me and everything he taught me. I certainly can write a 100 page paper on the things with which he tried to help me, all of which I fondly remember.

My thesis committee guided me and taught me all these years. Thank you to Dr. Barrett Caldwell, Dr. Ji Soo Yi, and Dr. Paul Lee. Dr. Caldwell taught me to look at the world and problems from a different perspective. He stopped me after class once and encouraged me with a sweet word, which helped me build self-confidence. Thank you. I want to thank Dr. Yi for his enthusiasm and love for teaching and research. It is definitely contagious. Thank you to my mentor, Dr. Paul Lee, at NASA Ames Research Center, whom I often refer to as “Uncle.”

I do not think I could have completed my Ph.D. without his help and support. Thank you so much for all the advice and guidance. Thank you for being my “Uncle Paul.” Thank you in so many different ways.

I would also like to express my sincere appreciation to Ms. Nancy Smith and Dr. Kevin Jordan for providing me with the great opportunity to work at NASA Ames Research Center. They have always been looking for ways to make my life easier, help me build self-confidence, and train me to become a researcher. Thank you for being great mentors. Also, thank you for the support in many ways. Again, I do not think I could complete my Ph.D. without your help.

I would like to thank my peers at Purdue University for the support. Special thanks go to Arik Q. Dao, Shan Xie, Julian Archer, and Nguyen VP Nguyen for spending all the fun late nights together in the lab at Purdue. I also want to thank my colleagues at the Airspace Operations Laboratory at Ames for their help and kindness. I want to thank Dr. Everett Palmer for mentoring me and showing me his enthusiasm for knowledge and research. I also want to thank Dr. Eric Chevalley and Nancy Bienert for bringing me great excitement and happiness to my life. It was always fun having countless lunches with them. I do appreciate their time. Thank you. I also want to thank Jeffrey Homola for his big smile. It was always cheerful. I also would like to thank Dr. Bonny Parke, Dr. Thomas Prevot, Dr. Todd Callantine, and Constantine Speridakos for their continuous support and help. Thank you.

Lastly, I want to thank the children I met during my mission trip to Cambodia. They have transformed my life, taught me how to appreciate things in my life. They taught me whatever I have is not a given. Thank you.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

LIST OF ABBREVIATIONS

COCOM: Contextual Control Model

EEG: Electroencephalogram signals

TRACON: Terminal Radar Approach Control

ETA: Estimated Times of Arrival

ACT-R: Adaptive Control of Thought

Midas: Man-Machine Integration Design and Analysis System

TP: True Positive

TN: True Negative

FP: False Positive

FN: False Negative

Hz: Hertz

RMSE: Root Mean Square Error

RMSD: Root mean square deviation

RT: response time

pre-PB: Before PB is subjectively identified

Post-PB: After PB is subjectively identified

ROC: Receiver Operating Characteristic

Bpm: Beats Per Minute

ABSTRACT

Yoo, Hyo-Sang. Ph.D., Purdue University, December 2015. Detection of Operator Performance Breakdown in a Multitask Environment. Major Professor: Steven Landry.

The purpose of this dissertation work is: 1) to empirically demonstrate an extreme human operator's state, performance breakdown (PB), and 2) to develop an objective method for detecting such a state. PB has been anecdotally described as a state where the human operator "loses control of the context" and "cannot maintain the required task performance." Preventing such a decline in performance could be important to assure the safety and reliability of human-integrated systems, and therefore PB could be useful as a point at which automation can be applied to support human performance. However, PB has never been scientifically defined or empirically demonstrated. Moreover, there exists no method for detecting such a state or the transition to that state. Therefore, after symbolically defining PB, an objective method of potentially identifying PB is proposed. Next, three human-in-the-loop studies were conducted to empirically demonstrate PB and to evaluate the proposed PB detection method.

Study 1 was conducted: 1) to demonstrate PB by increasing workload until the subject reports being in a state of PB, and 2) to identify possible parameters of the PB detection method for objectively identifying the subjectively-reported PB point, and determine if they are idiosyncratic. In the experiment, fifteen participants were asked to manage three concurrent tasks (one primary and two secondary tasks) for 18 minutes. The primary task's difficulty was manipulated over time to induce PB while the secondary tasks' difficulty remained static. Data on participants' task performance was collected. Three hypotheses were constructed: 1) increasing workload will induce subjectively-identified PB, 2) there exists criteria that identify the threshold parameters that best detect the performance characteristics that maps to the subjectively-identified PB point, and 3) the criteria for choosing the threshold parameters are consistent across individuals. The results show that increasing workload can induce subjectively-identified PB, although it might not be generalizable — 12 out of 15 participants declared PB. The PB detection method was applied on the performance data and the results showed PB can be identified using the method, particularly when the values of the parameters for the detection method were calibrated individually. Next, study 2 was conducted: 1) to repeat the demonstration of inducing PB, 2) to evaluate whether the threshold parameters established in study 1 for the PB detection method can be used in a subsequent study, or whether they have to be re-calibrated for each study, and 3) to examine whether a specific physiological measure (pulse rate) can be used to identify the subjectively-reported PB point. Study 2 was conducted in the same task environment (three concurrent tasks) as study 1. Three hypotheses were constructed: 1) increasing workload will induce subjectively-identified performance breakdown, 2) the threshold parameters established from study 2 will be the same as those from study 1 for all participants and will perform approximately as well or better, and 3) there exists criteria for choosing the threshold parameters that captures the characteristics at the subjectively-reported PB point using the PB detection method on pulse rate data. The results show that increasing workload induces the same participants (12 out of 15) from study 1 to declare PB. Also, it was found that the threshold parameters established in study 1 for the PB detection method cannot be reliably used in a subsequent study, and suggest that it may require re-calibration for each study. The results provided no evidence that pulse rate data can be used to detect PB. Study 3 was conducted: 1) to determine if PB is induced by the primary task workload or is affected by the presence of the secondary tasks, and 2) to re-test whether threshold parameters from study 1 can be used in a subsequent study. In study 3, the same participants from study 1 and 2 were only asked to perform the primary task while its difficulty increased in a similar manner to the first two studies. Two hypotheses were established: 1) PB will occur without the secondary tasks being present, and 2) the threshold parameters established from study 3 will be the same as those from study 1 and/or study 2 for all participants and will perform approximately as well or better. No participants declared PB without the secondary tasks present, even though the primary task workload was the same. Again, it was verified that the threshold parameters established in study 1 and 2 for the PB detection method cannot be used in a subsequent study, and suggest that it may require re-calibration for each study.

CHAPTER 1. INTRODUCTION

Anecdotally, most people are familiar with the sensation where, during a task with very high workload, a state is reached where the operator goes “hands off” and completely drops the task. Such an extreme state is referred to here as performance breakdown (PB). It is important to prevent such a state from being reached, particularly in a safety critical system that requires a human operator to assure the safety and reliability of the system’s operation. If PB can be detected in advance, then it may be prevented from occurring by allowing automation to intervene and assist or replace the human operator.

However, PB has been only anecdotally described in past research and has never been scientifically identified or empirically demonstrated in an experimental setting. This dissertation work contributes to filling those gaps. Moreover, this dissertation contains the initial ground work that could lead other researchers to extend the work and further explore PB and its detection method in the future.

The dissertation is organized in the following way: The second chapter presents: 1) past studies on PB (based on resource theory), 2) studies that characterized and classified different human performance characteristics, 3) categories of techniques that have been identified to monitor various changes in the state of the human operator, and 4) the system design concepts of function allocation systems to show where PB detection can be used to build an effective and safety-assured human integrated systems environment. The third chapter presents an objective method for identifying PB. Chapters four, five, and six describe three human-in-the-loop studies that were conducted to empirically characterize PB and demonstrate how well the objective PB identification method performs. The last chapter summarizes the results, discusses their implications and proposes further research opportunities.

CHAPTER 2. BACKGROUND

2.1 Performance Breakdown (PB) in Previous Studies

According to past studies, PB occurs when task demand exceeds resource capacity (Wickens, 2008). Also, it has been identified that workload is the primary source of resource depletion, and the scarcity in mental resources may be a cause of performance degradation (Khaneman, 1973).

2.1.1 Resource Theory

The limited-capacity resource model was first introduced by Kahneman (1973). He suggested that there is a limited pool of mental resources that can be allocated to tasks. There are two prominent perspectives on resource theory. One concept is the central resource theory and the other is multiple resource theory. Central resource theory suggests that there is a central reservoir of resources that can be allocated to complete tasks (Kahneman, 1973). Multiple resource theory posits that there are multiple pools of resources that can be utilized simultaneously (Wickens, 2002).

2.1.1.1 Central Resource Theory

This perspective of resource theory was introduced by Kahneman (1973). In the theory, it is proposed that two distinct tasks can be performed successfully and simultaneously as long as the resources required for performing both tasks do not take up the entire pool of resources. Kahneman (1973) theorizes that the central pool of resources varies according to the arousal of individuals. Maximum resources are thought to be available when a person's arousal level is at an optimal level for the situation.

In central resource Theory, the amount of arousal can be controlled by two sets of factors (Kahneman, 1973): 1) the task demand put on humans by engaging or preparing to engage in task activities and 2) other external factors independent of the task demand, such as the intensity of stimulation, the physiological effects of external factors (e.g., drugs), or the condition of the person (Gjerde, 1983). Task demand is defined as the amount of mental effort that is required to perform the task. Failure to provide an adequate level of effort results in performance degradation.

Regarding resource allocation for performing concurrent task management, attentional conflict is likely to be created when a person is demanded to perform competing tasks simultaneously (Hoffman, Nelson, & Houck, 1983). This conflict is a result of the dual demand on resource allocation. According to Kahneman's model, there are three rules governing an individual's resource allocation policy when performing multiple tasks. First, individuals like to ensure completion of at least one task. Second, based on the novelty and meaningfulness of the event, attentional resources are allocated involuntarily. Third, momentary intentions, which can be driven by enforcement from instructions, can drive resource allocation. This theory suggests that the capacity limits may impact human performance.

2.1.1.2 Multiple Resource Theory

Multiple resource theory represents separate and relatively independent processing. It is a theory of simultaneous multiple task performance, and can predict dual task interference levels between simultaneously performed, time-shared tasks in a multitask environment (Wickens, 2008). For example, the concept would expect independent processing between visual and auditory processing. The introduction of this concept allows variability in task interference (i.e., time-shared concurrent performance) to be explained better than the central resource theory (Kahneman, 1973), which posits that there is only a single pool of resources with a finite limit (Liu, 1997).

According to Wickens (1984), there are three possible factors (i.e., confusion, cooperation, and competition) that can be involved in concurrent task management performance outcomes (Wickens, 2008). The first is confusion of task elements, where performing similar tasks results in the tasks interfering with each other. The second is cooperation, where high similarity in tasks yields collective results. The last factor, resource competition, is where tasks compete for limited resources.

In multiple resource theory, there are three general components: 1) stages of processing (e.g., perception and central processing), 2) modalities of both input (e.g., visual and auditory) and output (e.g., manual and verbal), and 3) types of coding in memory (e.g., verbal and spatial). The stages of processing demonstrate how functionally separate resources are used for different stages of information processing. Figure 2 illustrates this for two different resources at two different stages.

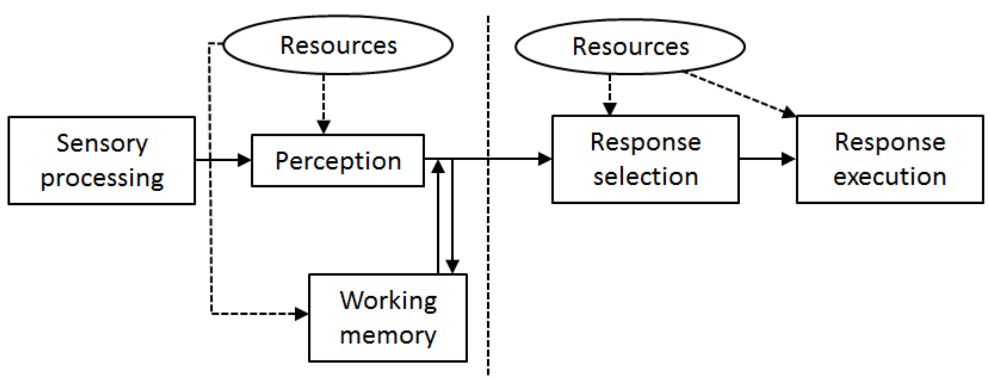


Figure 2.1 Representation of the usage of different resources at two stages (by Wickens, 2002)

The idea of two input modalities explains how performance can be better if task demands are shared among both modalities instead of using only one (Brown, 1997). In general, it is easier to divide attention into a visual and an auditory task than to perform two visual tasks alone or two auditory tasks alone, as there will be less interference between the modalities (Navon, & Miller, 1987; Brown, 1997; Wickens, 2002).

The distinction is also made between spatial and verbal resources. This separation of spatial and verbal resources seems to explain how manual (spatial) and verbal control are often effectively time-shared (Wickens, 2008). There have been several studies that support this distinction (Shah, & Miyake, 1996; Recarte, & Nunes, 2000). Multiple resource theory allows the prediction of changes in performance based on the characteristics of two or more time-shared tasks that are assigned to operators (Wickens, 1981).

According to the model, there is a capacity limit in the amount of resources one can utilize (Liu, 1997). Hence, if humans are required to perform multiple time-shared tasks that require the use of a common pool of resources, they may experience PB (Wickens, 2002).

2.1.1.3 Single Channel Bottleneck Theory/Filter Theory

Based on the Single Channel Bottleneck theory, there is a bottleneck in the human's central processor which imposes limitations on the ability to effectively perform multiple tasks concurrently (Craik, 1948; Liu, 1996; Liu, 1997). This fits in with the critical aspects of the filter theory first proposed by Broadbent (1958) that proposed that information can only be processed serially within a given structure. In this model, time is the only resource that is considered critical, and tasks can be queued up to be processed in a serial manner (Hendy et al., 1997).

This theory posits that the load on the human's processing of information is a direct result of the ratio of the time required to process the information to the total available time for making a decision (Hendy et al., 1997).

This approach describes performance as the ratio of information processed to information demanded, where the rate of information demanded is computed using the task difficulty and availability of time. The amount of information to process could vary significantly depending on the resource allocation strategy of the individual.

This theory has been criticized because people carry out information processing in parallel rather than in series. There are also cases in which task difficulty cannot be directly correlated with time demands (Carpenter et al, 1999). Operators can often perform two tasks with little or no interference effects on each other. Furthermore, the bottleneck in the central processor can be considered as the limited availability of all resources, rather than just attention limits.

2.1.2 Workload

Workload is thought of as a primary source of resource depletion (Kahneman, 2001) and it has been empirically demonstrated that workload drives changes in performance. Multiple definitions of workload currently exist. To some researchers, mental workload is conceived of as time pressure (Hendy at al., 2001). In another definition, mental workload is related to the cognitive resources required to perform a task in which the operator has become actively engaged (Gopher, & Donchin, 1986). In some sense, workload is considered to be the cost incurred by a human operator to achieve a particular level of performance (Andre, 2001).

Task load is often confused with workload. However, they should be distinguished, as several researchers have previously done. According to Hilburn and Jorna (2001), task load is the demand imposed by the task itself, and workload is the subjective experience of the task demand that the human perceives.

There are many ways to measure workload. Perhaps this should not be surprising, since there is no single definition of workload and therefore there should be no single metric to measure it.

Even if there were a single definition to describe all the cognitive dimensions associated with workload, measuring it would be difficult. In some circumstances, workload activates anxiety or frustration, and that might further interfere with performance (Dell’Erba, Pancheri, & Intreccialagli, 1988; Whinghter, Cunningham, Wang, & Burnfield, 2008). In addition, high and prolonged workload creates stress that is also known to increase fatigue (Parshuram, Dhanani, Kirsh, & Cox, 2004).

Since it is hard to isolate the cognitive factors associated with workload, what the experimenter observes and measures is not necessarily workload itself, but may be simply a change in whatever measurement the experimenter chooses to use.

An example that helps demonstrate the complexity of measuring workload is the study that has been conducted to empirically demonstrate the curvilinear relationship between arousal and performance (the inverted U-shape model). Yerkes and Dodson (1908) gave mice different intensities of electric shock to determine the effect on performance. The results indicated that mice perform better with higher intensity shocks. This became the Yerkes – Dodson Law of Performance. However, the conclusions are supported weakly and only in the context of the study that was conducted (Hancock & Ganey, 2003; Lagu, Landry & Yoo, 2013; Neiss, 1988).

Despite the variety of definitions of workload, a few factors consistently appear to influence what has been measured as workload from past studies, which are task difficulty and task load (Hancock, Williams, & Manning, 1995; Gevins, Smith, McEvoy, & Yu, 1997; Veltman, & Gailard, 1998), which will be tested in this dissertation as the contributing factors of PB.

2.2 Human Performance Behavior

This section discusses previous work on characterizing and classifying different types of human performance behavior. The two most prominent models of human performance behavior are presented: 1) the skill, rule, and knowledge based information processing model, and 2) the contextual control model (COCOM).

2.2.1 Skill, Rule, and Knowledge Based Information Processing Model

In the previous studies, different types of performance behaviors are classified into several categories. The most well-known classification is developed by Rasmussen (1979). In this framework, human information processing behaviors are categorized into three types: skill-, rule-, and knowledge-based information processing.

In the skill-based mode, the human possesses minimal conscious control in performing an action. This type of behavior generally represents the smooth execution that does not require much attention to perform tasks, which may result in the liberation of cognitive resources. This type of action often involves physical activities.

In the rule-based mode, the level of conscious control is at the intermediate level that is between the knowledge- and skill- based modes. The rules and procedures determine what actions should be taken in the next step, where the rules could be learned through experience, interaction, and training. In the process of performing tasks, the rules and procedures determine what actions should be taken in the next step.

In the knowledge-based mode, an advanced level of reasoning dictates the control actions, and these actions are made in a completely conscious manner. This mode demands more high-level cognitive processes than the other two modes, as it involves a thorough understanding of the situation and context to plan the actions required.

Although Rasmussen's classification of human behavior is well-known, it has been disputed for its simplicity. His classification does not sufficiently explain the flexibility and variety of human cognitive and motor skilled behaviors (Caldwell, 1997) including the condition where the human completely loses control of tasks (PB). In addition, this model does not consider the impact of the dynamic environment in which human behavior is studied (Hollnagel, 1993; Stanton, Ashleigh, Roberts, & Xu, 2001).

2.2.2 Contextual Control Model (COCOM)

The Contextual Control Model (COCOM) was introduced by Hollnagel (1993) to provide a useful framework that describes human performance as a set of a multiple functional control strategies and he has identified, particularly, different control mode of human relation to task environment. There have also been a number of models describing switch or selection processes of control strategies, which occur under the effect of external stressors (Maule, 1997; Todd, et al., 1995; Kerstholt, 1996).

In COCOM, Hollnagel (1993) states that “the degree of control a person will have over a situation can vary.” Control may also vary in a continuous form (Feigh, Pritchett, Jacko, & Denq; 2005). To describe this continuous form of control, Hollnagel has classified four contextual control modes (1993): scrambled control, opportunistic control, tactical control, and strategic control.

In scrambled control mode, the next action is irrationally or randomly chosen. Human performance is more likely to be trial-and-error. The human has no control and acts in an unplanned manner. An example of the context that may lead to this control mode is an emergency situation, where the subjectively determined available time is very limited and the situation is very unfamiliar.

In opportunistic control mode, the next action is determined by the salient features of the current context, which involves little strategic planning and anticipation. The operator’s perception of the available time may add constraints. Lack of knowledge and low familiarity are often the cause of this control mode (Hollnagel, 1993).

In tactical control mode, performance follows a known procedure or rule, which is externally driven or taught. This happens when there is a situation where a person’s event horizon goes beyond the dominant needs of the current situation but the immediate demands drive the next action. The operator's perception of the available time is considered to be limited but adequate in this context, and the task is considered to be somewhat familiar (Hollnagel, 1993).

In strategic control mode, the human looks ahead at higher-level goals. In this mode, performance is guided by plans based on the consideration of goals. Human performance is expected to be better in this control mode than the others. The operator's perception of available time is abundant and a person's familiarity to context is high (Hollnagel, 1993).

Recently, there have been efforts to empirically test these different modes and identify their performance characteristics (Stanton, Ashleigh, Roberts, & Xu, 2001; Feigh, Pritchett, Jacko, & Denq, 2005).

2.2.3 Five Categories of Techniques for Monitoring the Human Agent's State

Previous studies offer three different categories of techniques for monitoring the state of the human agent: 1) performance measure based, 2) physiological measure based, and 3) others (critical event based, model based, and hybrid techniques). This section describes the techniques in more detail to explore what could be done to monitor for PB.

2.2.3.1 Performance Measure Based Techniques

Performance measure based techniques (Parasuraman, Mouloua, & Molloy, 1996) directly evaluate the performance of the human operator to monitor changes in the human's state.

This approach assumes that performance measures are an indirect reflection of the human's changes in various cognitive states. Any indications such as significant increases or decreases in performance are a reflection of the human's state changes.

In past studies, one of the most commonly practiced approaches for monitoring state changes based on performance measures is the setting of a threshold value. For instance, in a study conducted by Parasuraman, Cosenzo, & De Visser (2009), the human operator's state is considered to change when the accumulation of errors exceeds 60%.

However, it is often not possible to collect sufficient performance measures during the operation to find indications of change in state, as there are systems that require very minimal input from the human operator to successfully operate the system (e.g., monitoring tasks).

2.2.3.2 Physiological Measure Based Techniques

Physiological measure based techniques (Prinzel, et al, 2000; Prinzel, et al 2003) can be used to overcome the limitations of the performance measure based technique by collecting data regardless of the human operator's frequency of input to the system. A key assumption of the physiological measure based techniques is that changes in selected measures (e.g. heart rate, EEG) directly reflect changes in the cognitive state of the operator.

Many physiological measures are currently utilized, such as heart rate variability, electroencephalogram (EEG) signals, and pulse rate. For instance, pulse rate is believed to be sensitive to changes in task difficulty or task load (Jorna, 1993; Roscoe, 1993; light, & Obrist, 1983; Wright, Contrada, & Patane, 1986; Backs et al., 2003; Chen et al., 2008; Haarman et al., 2009; Ting et al., 2010; Lagu, Landry, & Yoo; 2013). Heart rate variability (HRV) is another popular physiological measure that has been practiced. HRV is believed to be sensitive to changes in emotion or cognitive demand (Archarya, et al., 2006; Lewis, 2005).

In the past, many investigations were conducted to show which measures are more sensitive to particular changes in task or performance (Wilson, & Russell, 2003; Mikulka, Scerbo, & Freeman, 2002; Wilson, Caldwell, & Russell, 2007; Sauer, Kao, Wastell, & Nickel, 2011). Although it is found that a certain physiological measure may be more sensitive than the others in reflecting a specific dimension (e.g., workload, engagement, fatigue, etc.), there are no clear guidelines to which measure to use for monitoring which dimension of an operator's cognitive state. It is also unclear how changes in physiological measures map to various performance characteristics.

The most commonly practiced approach for detecting changes in state is to use a threshold approach (Lagu, Landry, & Yoo, 2013). In this approach, a threshold is established for each individual's physiological responses. When the measure goes below or above the threshold, then the change in the state is determined to have occurred (Bailey at el., 2006; Lagu, 2009).

In addition to the threshold approach, researchers recently started investigating a new method to detect changes in operator state. The selected physiological measure (e.g., heart rate, heart rate variability, or EEG) is clustered and mapped to pre-defined classifications of the state (Wilson, & Russell, 2003; Ting et al., 2010). One of the techniques is k-means clustering, which is a method of finding groups within a data set. Initially, k number of centroids are selected and all points in the data are assigned to the closest centroid while the centroid of each cluster continuously gets updated until it does not change (Kanugo et al, 2002). This technique can be performed to identify the relationships between two variables, which is performed in the latter part of this document.

2.2.3.3 Other Techniques

There are three other techniques to monitor a human operator's state changes, which are: critical event based, model based, and hybrid based techniques. In the critical event based techniques (Hilburn, Jorna, Byrne, & Parasuraman, 1997), occurrence of a critical event is used to foresee potential changes in contextual demands that are imposed on the human. For instance, in air traffic control operations, a metroplex has multiple airports with a high traffic demand (Saraf, Clarke, & McClain, 2010). In such airspace, the departure fixes are shared by several airports in the area which often feed high departure flows during the same time period (McClain et al, 2009). This is often called a “traffic jam” or “rush.” During this period, the number of departures that are expected to arrive at the departure fix may exceed what the Terminal Radar Approach Control (TRACON) controller can handle. This could be predicted based on the Estimated Times of Arrival (ETAs) at the fixes during this time. The prediction therefore could identify potential changes in the human’s state due to high workload that is foreseen in the near future (Farley, et al., 2001).

Another example of the critical event based monitoring approach is that the level of automation increases in an air defense system in response to events, i.e., the activation of a ‘pop-up’ weapon (Barnes & Grossman, 1985). However, this type of event-based automation is insensitive to the actual changes in the state of the human, as it predicts based only on the context changes in an environment where human operator is performing.

Model based techniques use a human performance modeling approach to predict the human's state changes. Currently, various types of models have been developed, including ACT-R, Air MIDAS, and Core MIDAS (Leiden et al, 2001). The general method used for human performance modeling is to rigorously determine the causes associated with the human's potential state changes. Hence, similar to critical event based techniques, the disadvantage of this modeling approach is that it is insensitive to actual changes in state.

The four monitoring techniques (i.e., performance, physiological, critical event, and modeling approach) just described could have complementary benefits, which suggests that a hybrid of the techniques may leverage the advantages of each (Parasuraman et al, 1992). Changes in performance measures are related to changes in the internal mental process or physical condition of the operator, as well as changes in situation/context. However, without a clear mapping between the internal and external characteristics of human behavior, deriving clear information is difficult. Hence, the critical event based prediction or modeling approach could be used to clarify and acquire information needed to predict or identify changes in the state of the human operator in addition to other measures.

2.3 Function Allocation

This section describes one area where the findings of this dissertation work can be applied. A better understanding of performance breakdown (PB) and its detection can be used to build a safer and more reliable human-integrated systems environment.

2.4 Static vs. Dynamic Function Allocation Systems

Function allocation is a concept for distributing functions between the human and automation to design an effectively operating human-integrated system. Function allocation systems can be divided into two general categories: static and dynamic (Landry, 2012). Static function allocation systems allow distribution of functions between the human and automation to be done only once prior to the system execution.

This type of function allocation system is known to have shortcomings. In this system, the allocation decisions are done independently of time and contextual changes and are determined primarily based on the subjectively-identified strengths and capabilities of the agents (Parasuraman & Riley, 1997; Parasuraman, Sheridan, & Wickens, 2000). However, such attributes of the agents are not necessarily fixed; they may vary in time. The limitations due to the static design approach's inflexibility have been found to result in an imbalanced workload and a lack of system awareness on the part of the operator (Byrne & Parasuraman, 1996; Kaber & Endsely, 2004; Rouse, 1988; Scerbo, 1996; Wilson & Russell, 2007). Hence, it has led researchers to investigate a new concept for allocating functions, which is dynamic function allocation (Caldwell and Onken, 2011; Inagaki, 2003; Landry, 2012).

The design principle of dynamic function allocation is to allow functions to be dynamically allocated between agents to regulate the fluctuation of the human's state, which continuously changes during operations. For example, when there is a significant increase in air traffic density in an en route sector, an air traffic controller may not be able to effectively manage such high traffic density. Erzberger (2004) empirically demonstrated that controllers can manage only approximately 15 aircraft in their sector at any one time. Past research has shown that, when the traffic exceeds 1.5 times that level, it becomes unmanageable (Prevot, Homola, & Mercer, 2008). In such traffic, the controller's workload reaches its maximum threshold, where human's state is no longer at the desirable level. Such an increase in workload could be used as a predictor for changes in operator state and could be used to allow automation to intervene to assist with the controller's tasks. Figure 2.2 demonstrates the dynamic change in contextual demand in air traffic management that causes an imbalance in workload during operations.

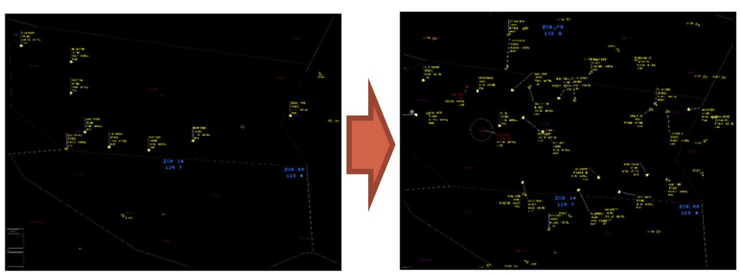


Figure 2.2 Increase in air traffic density shown on an air traffic controller's scope

2.4.1 Adaptable vs. Adaptive Automation Systems

Dynamic function allocation could be implemented in two different forms to improve human-automation interaction: Adaptive and adaptable automation

(Landry, 2012). The major distinction between the two forms is the possession of the decision authority for reallocating functions dynamically (Landry, 2012). In the adaptable automation form, dynamic reallocation of functions is done by the human agent. In the adaptive automation form, the system possesses the decision authority to dynamically reallocate functions among the agents based on continuous evaluation of the context and the state of the agent (Parasuraman, Bahri, Deaton, Morrison, & Barnes, 1992).

Several studies have investigated the efficacy of adaptable automation systems. In operations involving adaptable automation, the human agent possesses the final control authority and this results in better mode awareness, higher levels of task engagement, and increased user acceptance (Li, 2013; Miller & Parasuraman, 2007).

A number of studies have examined the potential benefits of adaptive automation systems (Parasuraman, Mouloua, & Hilburn, 1999; Hilburn, Jorna, Byrn, & Parasuraman, 1997; Kaber & Endsley 1997; Wilson & Russell, 2007). These studies have demonstrated that adaptive automation reduces workload compared to the traditional static type of function allocation (Parasuraman et al, 2009). Kaber et al. (2006) have empirically shown the potential improvement in situation awareness by implementing adaptive automation by running a human-in-the loop study.

Even though the potential benefits of adaptive automation systems have been demonstrated, there remain several issues regarding the building of an effective operational adaptive automation system. One of the major limitations comes from lack of understanding of determining when to allocate functions between the human and automation (Inagaki, 2003; Yoo, 2012).

The past work has focused on detecting subtle changes in workload or operator functional state, which is an estimate of how well human can perform tasks. The reliable detection of such changes can be used as an automation triggering point (Yoo, 2012). However, such subtle changes may be difficult to detect. PB is an extreme state of human operator, which should be more obvious to detect than such subtle changes. The focus of this dissertation work is on empirically demonstrating PB and developing an objective detection method for detecting/predicting PB. The findings of this work can contribute to extend our understanding of potential use of detection of operator’s state change as an automation triggering mechanism.

CHAPTER 3. PERFORMANCE BREAKDOWN (PB)

3.1 A Proposed Objective Way of Detecting Performance Breakdown (PB)

The PB detection method described in this dissertation distinguishes data into a binary form (PB vs. Non-PB) by setting the threshold on the selected measure for monitoring human operator's sate changes. The following describes the method in more detail, which could be extended as a framework for detecting transition in other cognitive states as well.

PB occurs when the human operator fails to maintain minimally acceptable performance in a primary task for some minimum duration or longer.

$$(p < p_{crit}) \cap (\Delta t > \varepsilon) \tag{3.1}$$

In the equation above (3.1), p refers to the human operator's performance on a specific task. p_{crit} is a minimally acceptable performance level for the task. $\mathcal{E}$ indicates a maximum duration of time allowed for adjusting performance to maintain performance above the minimum performance level (p_{crit}). Δt is the contiguous duration of time that an operator fails to maintain the minimum performance level (p_{crit}). Parameters ($p_{crit}, \mathcal{E}$) are most likely task specific, and may need to defined by subject matter experts or be empirically determined.

Performance (p_{crit}) can also be computed as an error rate, i.e. the number of correct or incorrect responses during a fixed duration of time. In such cases, Equation 3.1 can be modified accordingly. For example, the operator is asked to respond to twenty stimuli that are presented every two minutes. The total duration of the operation is thirty minutes. The operator's performance can be evaluated for every two-minute period by computing the error rate during that period. If the error rate exceeds the critical threshold value for an indicated duration of time, then PB is said to occur for that time period.

Performance can be computed as an error as well. For instance, the compliance of a pilot with a specified flight path could be considered the pilot's performance. In such a case, PB would be indicated if the pilot failed to keep the aircraft on the target route beyond the minimally acceptable deviation it for a minimum period of time.

Previous work has indicated the potential sensitivity issues associating with using the threshold approach for detecting changes in the human's state (Lagu, Landry, &, Yoo, 2013; Yoo, 2012). Hence, three evaluation criteria are identified, which can be used to evaluate the efficacy of parameters (p_{crit}, ε) in detecting PB. The three evaluation criteria are: sensitivity, specificity, and delay time to detection. These criteria are commonly used parameters in signal detection analysis (Bradley, 1997; Parasuraman, Sheridan, & Wickens, 2000; Kuchar, 1996).

The sensitivity was computed using the following equation (Swets, 2012; Swets, 2014):

$$Sensitivity = \frac{Total\ duration\ of\ true\ positive}{Total\ duration\ of\ true\ positive + Total\ duration\ of\ false\ negative} \quad (3.2)$$

In the equation above (3.2), the total duration of true positive (TP) indicates the time period that PB is correctly diagnosed as PB. The total duration of false negative (FN) represents the period when PB is incorrectly identified as not being PB (Non-PB).

The specificity was calculated using the following equation (Swets, 2012; Swets, 2014):

$$Specificity = \frac{Total\ duration\ of\ true\ negative}{Total\ duration\ of\ true\ negative + Total\ duration\ of\ false\ positive} \quad (3.3)$$

In Equation 3.3, the total duration of true negative (TN) is the period that the Non-PB condition is correctly identified as Non-PB. The total duration of false positive (FP) is the period when Non-PB is incorrectly identified as PB.

The following figure depicts an example of the false negative situation.

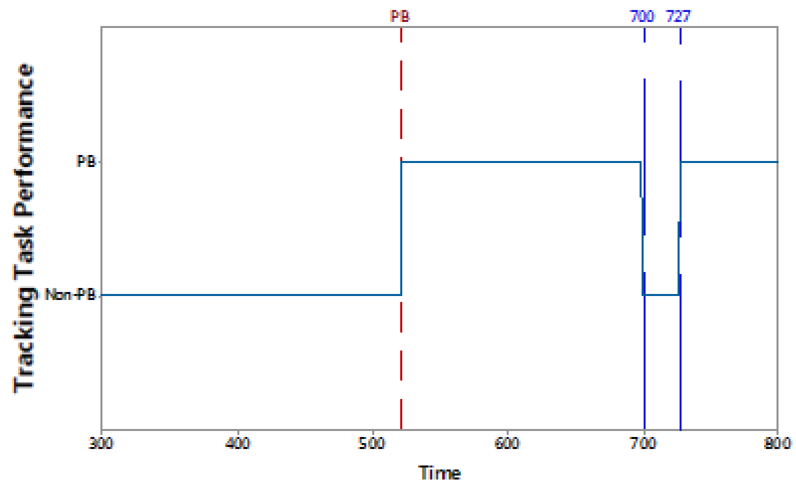


Figure 3.1 Nominal example of false negative

Figure 3.1 depicts a nominal example of the false negative situation. In Figure 3.1, a tracking task with increasing task performance over time results in PB, shown as the red dotted line after 500 seconds. Once PB occurs in a task with increasing task difficulty, it should continue as long as no resolution action is made. However, from 700 seconds to 727 seconds, it is identified that there is Non-PB. During the period, PB is incorrectly identified as not being PB (Non-PB) and produces FN.

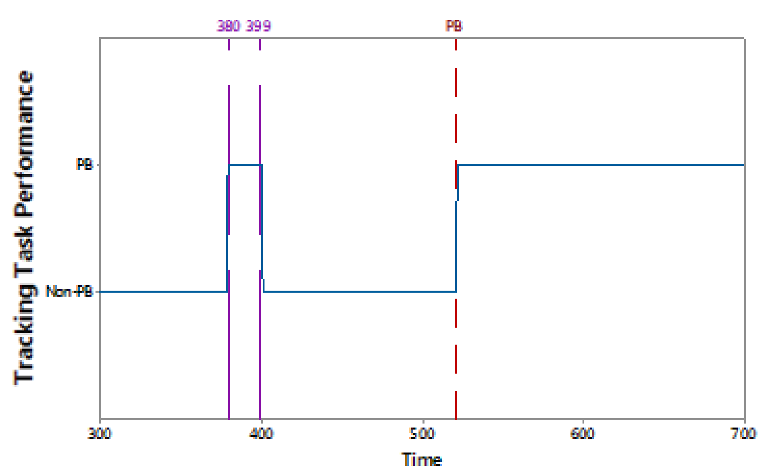


Figure 3.2 Nominal example of false positive

An example of a false positive is presented in the figure below (Figure 3.2). PB is shown to occur after 500 seconds. However, there is a time period (from 380 seconds to 399 seconds) that is identified as PB. During the period, the detection method incorrectly identified Non-PB as PB and produces false positive.

The delay time to detection is the period of time it takes from the point when PB occurs to the time the PB detection method detects PB. Having a large value for $\mathcal{E}$ is one of the major contributors for having a large delay time. When it is ambiguous to determine which combination of the parameters ($p_{crit}, \mathcal{E}$) work the best in detecting PB, the delay time could be used to as additional information that could determine which combination of the parameters is more effective in detecting PB.

A Receiver Operating Characteristic (ROC) curve can be constructed to investigate how various threshold values affect PB detection. The ROC curve helps determine the optimal threshold values that effectively balance the specificity and sensitivity (Bradley, 1997). Figure 3.3 shows an example of ROC curve. In the figure (Figure 3.3), the numbers in the right upper corner indicate different values for $RMSD_{crit}$ and the number on top of each dot in the graph represents the value that has been tested for ε.

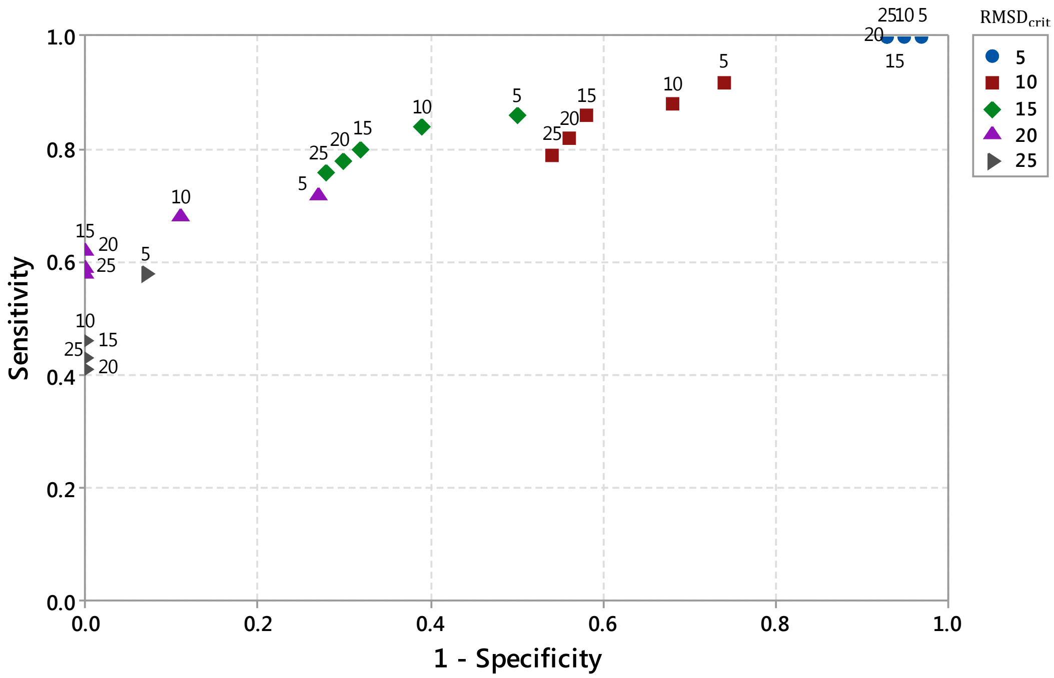


Figure 3.3 ROC curve: evaluation of parameters ($RMSD_{crit}$, ε) on participant's tracking task performance

In the ROC curve graph (Figure 3.3), the threshold values found with the shortest Euclidian distance to the left upper corner are sought to balance the competing characteristics the best (i.e., both maximizes the sensitivity and the specificity, assuming an equal cost to a false positive and false negative), which is referred to as criterion 1 in this dissertation work. This could be applied in the system where the false detection and missed detection are equally important. In the figure (Figure 3.3) above, $RMSD_{crit}$ = 15, ε =10 are identified based on criterion 1.

The combination of the threshold values that detect PB more conservatively can be also selected. The condition that shows the maximum specificity but had the highest sensitivity will be referred to as criterion 2 for the rest of the paper. This criterion could be applied to the situation where the impact of the missed detection is critical.

In the Figure 3.3, $RMSD_{crit}$= 15, ε = 20 satisfy such criteria.

CHAPTER 4. STUDY ONE

4.1 Overview

This chapter presents study 1 in detail. The focus of study 1 was: 1) to demonstrate PB by increasing workload until the subject reports being in a state of PB, and 2) to determine the parameters (p_{crit}, $\mathcal{E}$) for objectively identifying the performance characteristics that maps to the subjectively-identified PB point and determine if they are idiosyncratic. In order to induce PB in a multi-task environment, the three tasks (one primary and two secondary tasks) were given to the 15 (13 male + 2 female) participants to perform for 18 minutes. During the run, the difficulty of the primary task was increased every 2 minutes while the difficulty of the secondary tasks was maintained at a static level.

The participants were asked to verbally indicate when they experienced PB, and the times at which the participants declared PB were recorded. Even after identifying PB, the participants were asked to continue performing the tasks to the best of their ability. This was to mimic a real situation, where pilots will not completely give up on the tasks that are given to them, even if they are experiencing PB.

4.2 Method

4.2.1 Participants

There were a total of 15 participants (13 male + 2 female) in this study. The age range of the participants was 23 – 34 years old. The participants had no prior experiencing of performing the tasks. Participants with any types of disability that might prohibit them from performing the tasks were excluded, such as hand tremor, blindness, etc.

In this type of study, it was difficult to determine a sufficient sample size. The number of participants for this experiment was determined to detect one standard deviation with $p = 0.05$ for Type I error and $p = 0.20$ for type II error with the assumption that the tracking task error data will follow a normal distribution. It requires nine participants to detect such a difference. For the studies, I conducted studies on 15 participants.

4.2.2 Experimental Apparatus

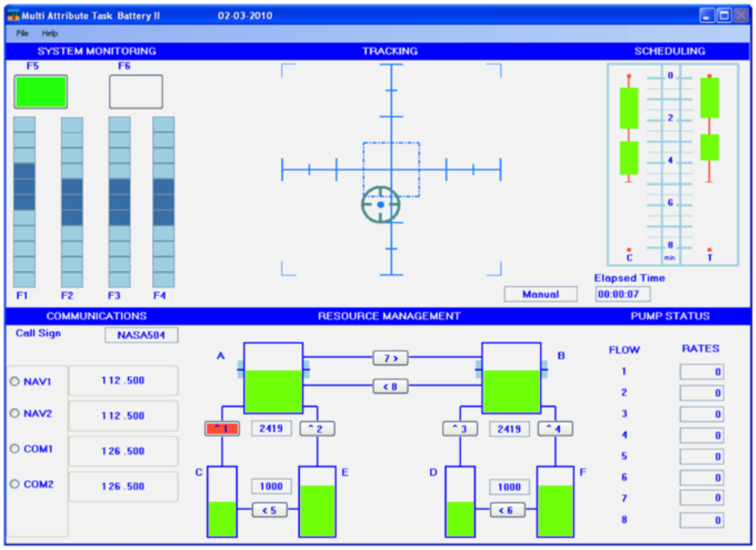


Figure 4.1 Screen shot of Multi-Attribute Tasks Battery II (MATB-II)

The study required participants to perform three tasks concurrently, which were the system monitoring task, the resource management task, and the tracking task from the latest version of Multi-Attribute Task Battery-II (MATB; Comstock, & Argegard, 1992). These tasks are designed in a way that mimics the general operations of a pilot's tasks in the cockpit environment. The primary task was a tracking task. The secondary tasks were system monitoring and resource management tasks. Instruction was provided to place strong emphasis on the primary task. The participants were provided with a keyboard and a joystick to perform the tasks.

These selected three tasks required perceptual attention, which theoretically utilizes the same non-sharable resources. The rationale for having only the tasks that share the same resources (i.e., these tasks are using only visual modality, manual output, and spatial coding) was to minimize the interference effect of using different pools of resources, such as the auditory modality, and to obtain the clear impact of changing tracking task difficulty on performance.

4.2.3 Independent Variables

In this study, there were nine (3 X 3) different levels of difficulty of the primary task that increased in steps to induce PB. The task difficulty was determined by the combination of two parameters: 1) the target movement, and 2) the joystick response sensitivity level. The target update rate varied based on the amount of random target movement per update cycle and the joystick response sensitivity levels varied based on the amount of influence the joystick movement has on target movement per update cycle.

Table 4.1 shows the nine conditions that were created to induce a step-wise increase in task difficulty. It was determined that high response sensitivity requires more effort than the medium or low level for the participants, as they tend to overshoot. It was determined that the medium sensitivity level provided the most comfortable sensitivity out of the three levels for the participants. Task difficulty was designed to increase every two minutes to provide sufficient time for the participants to realize the change in task difficulty.

Table 4.1 The combinations (3 x 3) of the target update rate and the response sensitivity levels

Task difficulty level	Target update rate	Response sensitivity
1	Low	Medium
2	Low	Low
3	Low	High
4	Medium	Medium
5	Medium	Low
6	Medium	High
7	High	Medium
8	High	Low
9	High	High

Each update cycle of the tracking task is 100 ms (i.e., 10 Hz). Figure 4.2 shows all possible directions for the next movement of the target in the tracking task. The target always starts at the center position (5). At every update cycle, the current position of the target is evaluated and random numbers are generated to determine whether to stay at the current position or to move towards one of the other states.

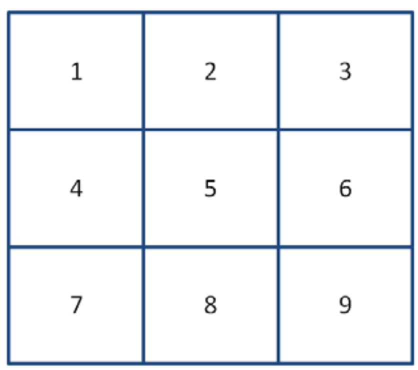


Figure 4.2 The target states of the tracking task

There are three different levels of target movement. At the low level, the rate moves only within one pixel unit from the current position. At the medium level, it moves two pixel units. At the highest level, it moves three pixel units. The tracking task display can be considered a regular quadrilateral-shaped unit grid with the origin in the center. The target can move left or right along the x-axis and up or down along the y-axis.

The sensitivity of the joystick response could be manipulated at three different levels (low, medium, and high). The value returned for the x-axis becomes greater as the joystick moves to the right side, and the value for the y-axis is greater as the joystick moves toward the user. The current and the last positions are evaluated every 10 Hz to compute the direction and speed of joystick motion. The sensitivity (i.e. the speed of the joystick motion) increases by 0%, 100%, and 200% respectively.

In addition to the difficulty changes in the primary task, there were two secondary tasks. The intention of having the secondary tasks was to allow the total task demand to be maintained at a high level so that the effect of the task difficulty changes in the primary task became visible.

4.2.4 Dependent Variables

There were three dependent variables: 1) time of PB that the participant verbally indicated, 2) root mean square error (RMSE) of the Primary task (pixel unit), and 3) errors in the secondary tasks – the resource management task and system monitoring task.

During the experiment run, the participants were asked to subjectively identify the PB point, and that time was recorded.

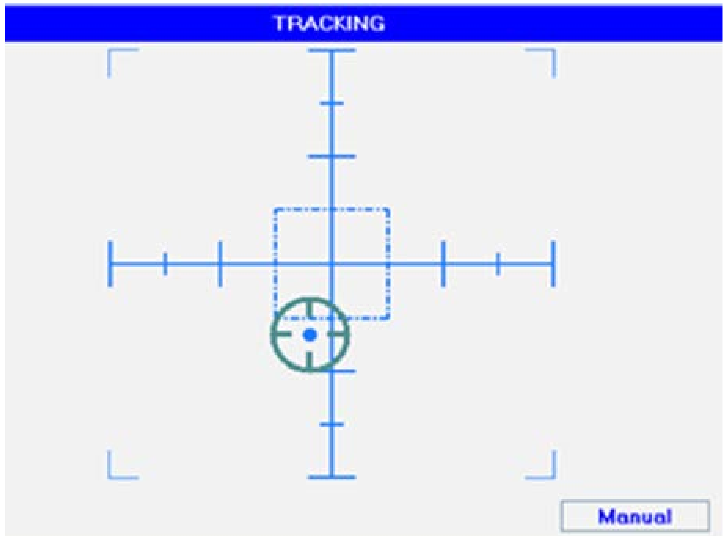


Figure 4.3 The tracking task

In the tracking task, the target continuously deviated from the center point. The participants' goal was to keep the target at the center point. The target positions were sampled twenty times per second and root mean square deviation (RMSD) values were recorded at every one-second interval.

$$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(0 - x_i)^2 + (0 - y_i)^2} \tag{4.1}$$

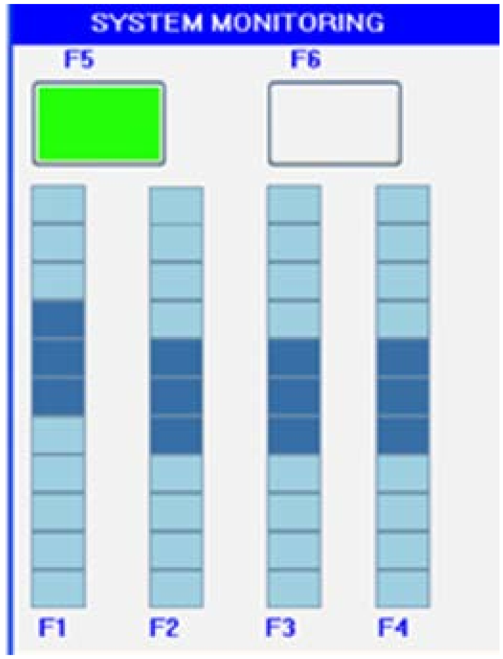


Figure 4.4 The system monitoring task

The system monitoring task, one of the secondary tasks, required the operator to monitor and respond to simulated warning lights and gauges. The minimum response time was set for all stimuli in this task. If participants failed to respond within five seconds, each failure was counted as an error. The participants were required to respond by pressing the corresponding function key. Both response time (RT) and the number of errors were recorded. An equal number of stimuli (a total of sixteen stimuli) were presented at random points within every 2–minute period.

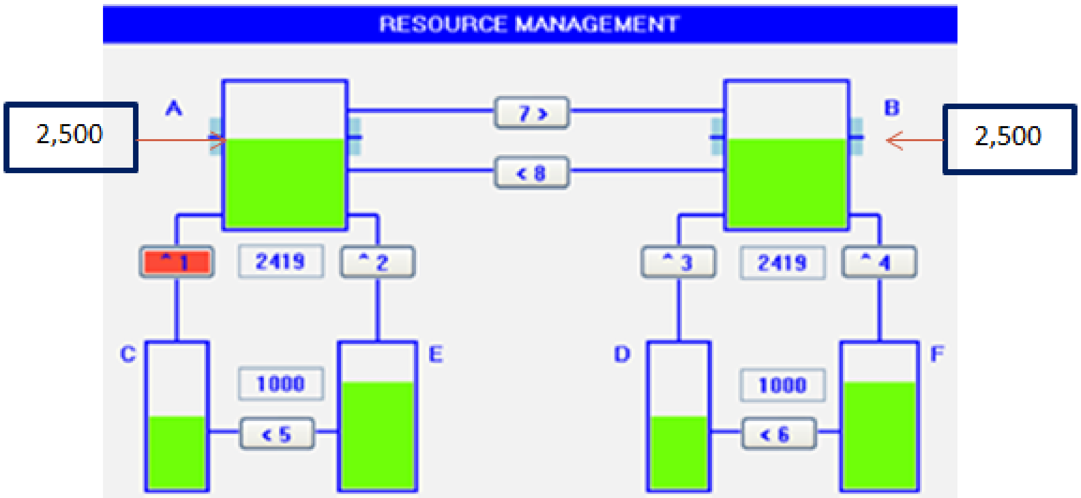


Figure 4.5 The resource management task

In the other secondary task, i.e. the resource management task, fuel levels in two primary tanks (A & B) had to be maintained at a target level (2,500 units). Deviations from the target level were recorded every ten seconds. The sum of absolute deviation from the target level in both tanks A and B were computed for the analysis.

4.3 Hypotheses

First, the following hypothesis was examined to determine whether an increase in workload induces PB.

Hypothesis 1a: Increasing workload will induce subjectively-identified PB.

As mentioned earlier, the PB detection method is task-specific, as the performance metric depends on the type of the tasks. Since the primary tracking task performance has been defined as RMSD error, the PB equation is as follows:

$$(RMSD > RMSD_{crit}) \cap (\Delta t \ \geq \varepsilon) \tag{4.2}$$

Equation 4.2 indicates that PB is identified when the deviation (RMSD) of the target for the tracking task exceeds the minimally acceptable performance level ($RMSD_{crit}$) for longer than a specified duration (ε). 5, 10, 15, 20, and 25 seconds were used as the values of each parameter ($RMSD_{crit}, \varepsilon$). The following hypothesis was constructed to test whether there exist a reliable criterion for choosing the combination of the parameters that identifies the subjectively-identified PB point.

Hypothesis 1b: There exists a criterion or criteria ($RMSD_{crit}$ and ε) such that the point in time corresponding to $(P < P_{crit}) \cap (\Delta t \ \geq \varepsilon)$ matches the subjectively-identified performance breakdown point.

Next, the following hypothesis was constructed to identify whether the criteria that were found to detect the subjectively-identified PB point is consistent across participants.

Hypothesis 1c: The criteria from hypothesis 1b are consistent across individuals.

4.4 Results

4.4.1 Overview

This section presents the results of the data analysis for study 1. The following are the summary results of the hypothesis testing: 1) increasing workload can induce subjectively-identified PB, although it might not be generalizable; 2) there were criteria that exhibit good performance in detecting the subjectively-identified PB point; however, 3) there were no such criteria that were consistent among participants.

In the following results section, the demonstration of how the detection method was applied is reported.

4.4.2 Hypothesis 1a

A total of 12 (10 male + 2 female) participants who indicated that they experienced PB, which supports the hypothesis 1a. (See Table 4.2.)

First, the histogram of the tracking task (RMSD) is visually observed.

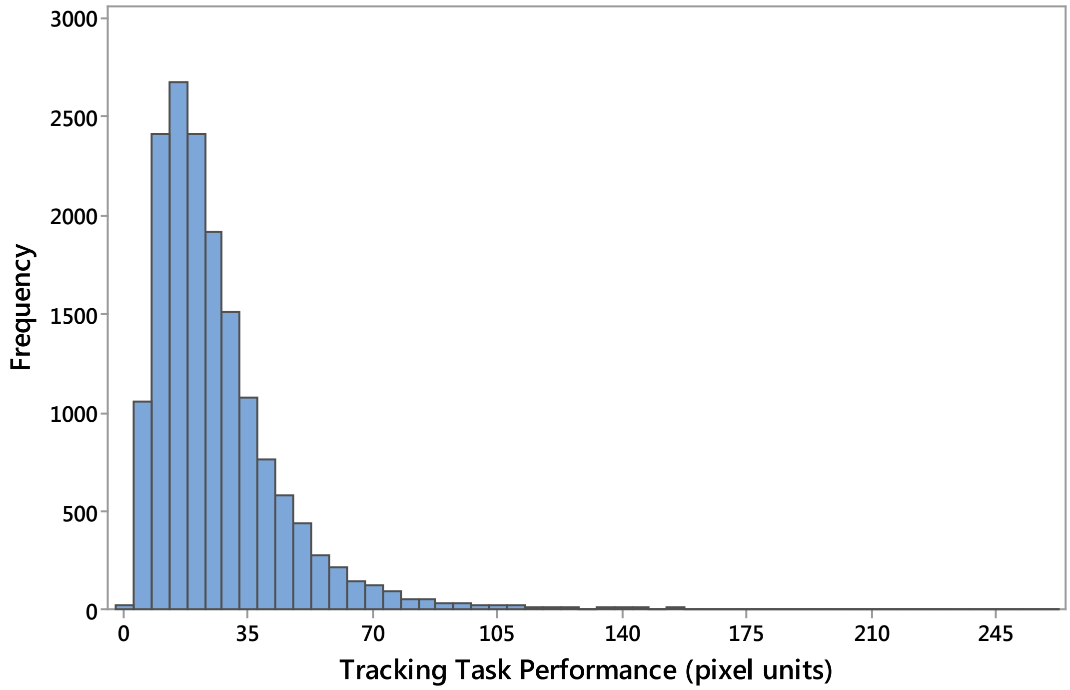


Figure 4.6 Histogram of the tracking task performance

In Figure 4.6, it is observed that the distribution of the tracking task performance has a left skew with a long right tail. Furthermore, the results of the normality test using the Anderson-Darling test indicated that the tracking task performance data is not normally distributed (see Figure 4.7).

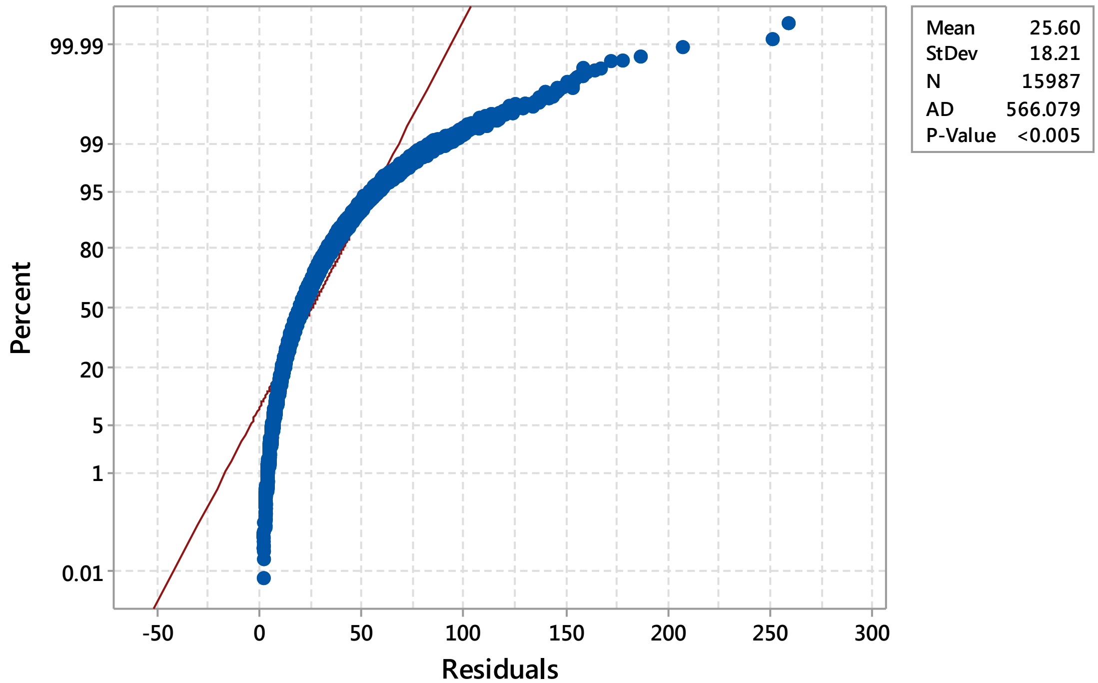


Figure 4.7 Anderson-Darling test on the tracking task performance

Table 4.2 Summary statistics of participants' tracking task performance (pixel units) and PB (PB = Yes or Non-PB = No)

Participant	Mean	SD	Median	PB
1	26.8	18.4	22.5	Yes
2	18.4	11.7	15.6	Yes
3	29.1	19.3	24.6	No
4	22.7	12.5	20.7	Yes
5	25.8	22.1	20.6	Yes
6	33.5	21.8	28.5	Yes
7	28.3	18.8	23.8	Yes
8	19.5	10.5	17.4	No
9	19.7	11.7	17.4	No
10	40.4	28.2	33.8	Yes
11	22.3	13.4	18.9	Yes
12	22.9	15.7	18.9	Yes
13	23.3	15.0	20.2	Yes
14	24.9	15.3	21.8	Yes
15	26.4	16.5	23.4	Yes

In Table 4.2, it is identified that there are individual differences in how the participants performed the tracking task (RMSD).

Figure 4.3 represents the raw tracking task performance of the participants, where a gradual performance degradation is observed. The raw tracking task performance of each participant with PB point indication are presented in Appendix A. Also Figure 4.8 shows the tracking task performance of all participants.

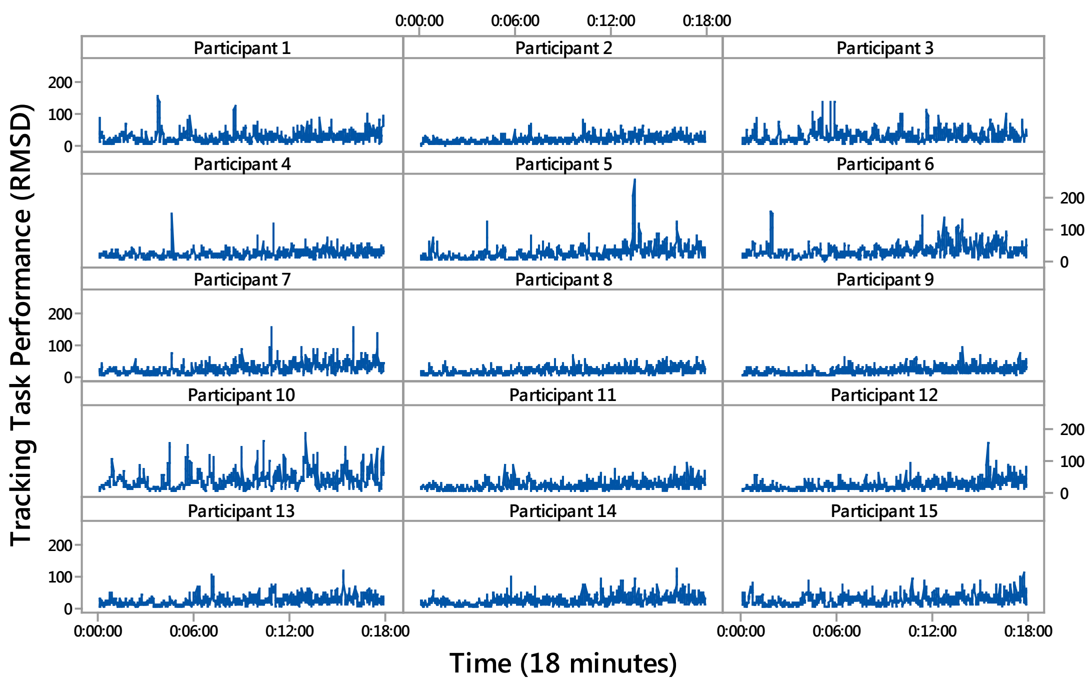


Figure 4.8 Participant tracking task performance (pixel units) vs. time

Table 4.3 presents the probability of the target moving toward the center point before and after PB for all participants. A proportion test was conducted to compare the probability of pre-PB with post-PB.

Table 4.3 Probability of target moving toward the center point

Participant	pre-PB vs. post-PB	Probability of target moving toward the center point	p-value
1	pre-PB	0.46	0.538
	post-PB	0.48	
2	pre-PB	0.44	0.142
	post-PB	0.49	
4	pre-PB	0.49	0.505
	post-PB	0.47	
5	pre-PB	0.44	0.181
	post-PB	0.49	
6	pre-PB	0.48	0.576
	post-PB	0.46	
7	pre-PB	0.44	0.359
	post-PB	0.47	
10	pre-PB	0.49	0.124
	post-PB	0.44	
11	pre-PB	0.48	0.447
	post-PB	0.5	
12	pre-PB	0.48	0.545
	post-PB	0.51	
13	pre-PB	0.47	0.681
	post-PB	0.45	
14	pre-PB	0.44	0.235
	post-PB	0.48	
15	pre-PB	0.46	0.827
	post-PB	0.45	

The task is designed in a way that the target moves toward the center point with a probability of 0.33 (=3/9) at every update cycle when there is no input from the participant. The probability that it stays at the same location is 0.11 (=1/9). The probability that it moves away from the center point is 0.55 (=5/9). The probability of the target moving toward the center point shows that the participants still provided inputs to maintain the performance of the tracking task even after PB, as there was no significant difference between before and after PB in the probability of correct target movement (see Table 4.3).

Then, the number of cases that the target moved toward the center point was identified and participants' performance on correcting the deviation was estimated. It is found that before PB, the magnitude of the target's movement toward the center point was greater than after PB (see Table 4.4). Given the probability of the target movements had no difference between before and after PB, such difference in the tracking task performance between before and after PB could be due to the participants' tracking task performance.

Table 4.4 The average magnitude of correction made on the target deviation from the center point

Participant	pre-PB vs. post-PB	The average magnitude of target moving toward the center point	p-value
1	pre-PB	7.2	0.0000
	post-PB	10.0	
2	pre-PB	4.2	0.0000
	post-PB	8.1	
4	pre-PB	4.5	0.0000
	post-PB	8.1	
5	pre-PB	5.9	0.0000
	post-PB	9.8	
6	pre-PB	7.9	0.0001
	post-PB	12.0	
7	pre-PB	4.9	0.0000
	post-PB	9.9	
10	pre-PB	6.5	0.0000
	post-PB	13.1	
11	pre-PB	6.2	0.0000
	post-PB	10.1	
12	pre-PB	7.6	0.0003
	post-PB	13.0	
13	pre-PB	4.6	0.0000
	post-PB	8.7	
14	pre-PB	6.1	0.0000
	post-PB	9.8	
15	pre-PB	5.6	0.0000
	post-PB	10.2	

The magnitude of the correction the participants were making was higher after PB. In addition, there was no significant change in how frequently participants provided inputs to the tracking task after PB. However, the average RMSD continued to increase throughout the operations. This shows that the participants still failed to maintain good performance for the tracking task. It suggests that the parameters ($RMSD_{crit}$, ε) in the detection method could be applied to the tracking task to capture PB.

4.4.3 Hypothesis 1b

The detection method has been applied to determine whether such types of degradation in performance can be sensitively captured by the PB detection method. A ROC curve was constructed (Figure 4.9) for each participant individually to investigate how various threshold values affect PB detection.

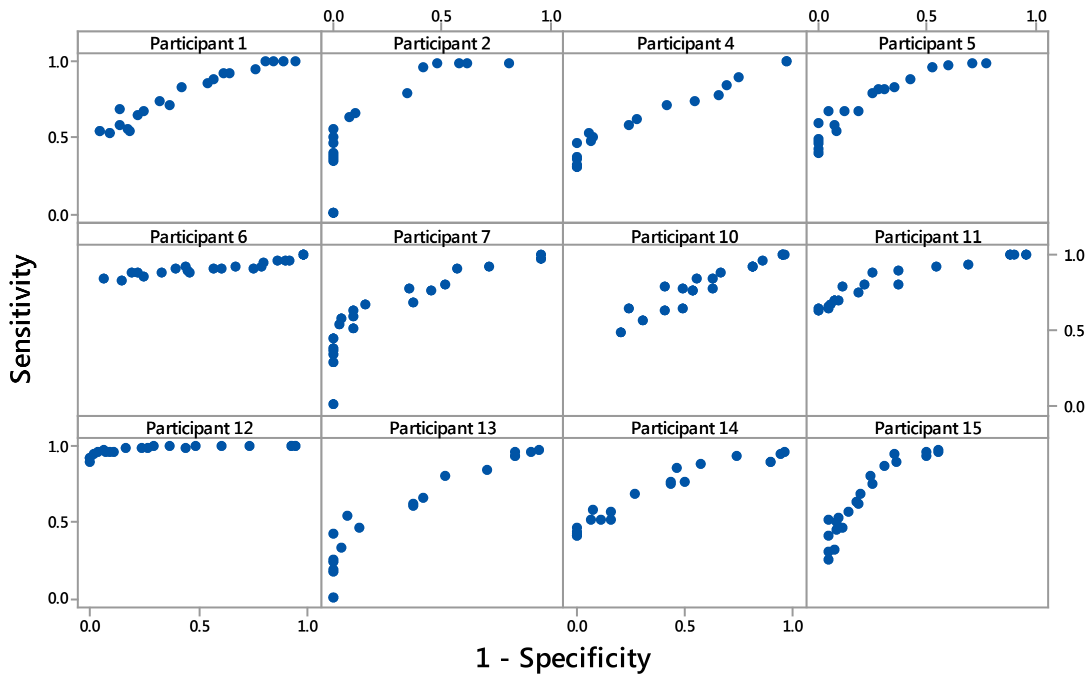


Figure 4.9 Receiver operating characteristic (ROC) curve: evaluation of parameters ($RMSD_{crit}$, ε) on PB detection

The effect of the different parameters ($RMSD_{crit}$, ε) on duration of false detection, missed detection, and delay time to PB detection on participants' tracking task performance was also determined. It was identified that the duration of false detection is inversely related to the value of $RMSD_{crit}$. It is also found that as the value of ε

increases, the false detection rate decreases. It is also found that the duration of missed detection of PB increases as the values of $RMSD_{crit}$ and ε increase (See Appendix A.) Although it is not perfect, there is a criterion that performs better in terms of detecting the subjectively-identified PB point.

4.4.4 Hypothesis 1c

Further investigation was conducted to determine whether there is consistency in the criterion among the participants. Table 4.5 reports how the average duration of false detection, missed detection, delay time, sensitivity, and specificity changes due to use of the different threshold values. According to the results shown in Table 4.5, it was found that there is no unambiguous criterion for choosing the optimal threshold values that perform consistently among the participants.

Table 4.5 The different combination of threshold values and their performance in average

$RMSD_{crit}$ (pixel units)	ε (seconds)	FALSE (seconds)	MISS (seconds)	DELAY (seconds)	1-Specificity	Sensitivity
5	5	448.8	4.7	0	0.9	0.99
5	10	436.5	6	2	0.87	0.99
5	15	422.3	7.2	2.5	0.84	0.99
5	20	413.1	12.7	3.3	0.81	0.98
5	25	403.7	12.7	4.2	0.79	0.98
10	5	342.9	68.6	14.6	0.67	0.99
10	10	292.2	101	7.5	0.56	0.89
10	15	254.9	140.5	10.7	0.48	0.85
10	20	219.5	180.7	19.2	0.42	0.81
10	25	134.7	216.9	37.2	0.36	0.77
15	5	226.8	179.8	3.4	0.42	0.81
15	10	151.5	261.2	15.3	0.27	0.73
15	15	103.6	338	79.8	0.19	0.67
15	20	78.4	396.2	65.7	0.19	0.62
15	25	65.7	422.3	192.5	0.11	0.58
20	5	145.7	295.8	10.3	0.26	0.7
20	10	79.1	390.8	86.4	0.14	0.62
20	15	56.6	509.8	172.1	0.1	0.51
20	20	35.6	551.1	105.8	0.06	0.49
20	25	12.3	570.9	187.2	0.05	0.46
25	5	84.9	395.5	21.8	0.14	0.62
25	10	42	512.2	112.7	0.08	0.52
25	15	22	498.9	201.7	0.04	0.45
25	20	13.2	517.1	282.9	0.03	0.44
25	25	12.3	529.3	140.7	0.02	0.41

The identified threshold values based on criterion 1 are presented in Table 4.8. In Table 4.6, it can be determined that there is no consistency among the participants in terms of which threshold values were found to meet criterion 1.

Table 4.6 The values of the parameters ($RMSD_{crit}, \varepsilon$) selected based on criterion 1 for each participant (study 1)

Partici -pant	$RMSD_{crit}$ (pixel units)	ε (sec)	Duration of False detection (sec)	Duration of Missed detection (sec)	Delay time to detection (sec)	1-Specificity	Sensitivity
1	15	10	197.1	146.6	0.0	0.3	0.8
2	10	10	55.9	260.4	45.1	0.2	0.8
3				No report of PB			
4	10	25	91.1	365.0	74.0	0.3	0.6
5	15	10	47.1	186.2	5.0	0.1	0.8
6	25	25	32.0	158.1	87.0	≈ 0.0	0.9
7	15	10	62.1	281.5	0.0	0.2	0.7
8				No report of PB			
9				No report of PB			
10	25	10	116.9	269.0	0.0	0.3	0.7
11	10	25	125.5	95.9	0.0	0.2	0.9
12	20	15	26.0	30.0	8.0	≈ 0.0	1.0
13	10	15	67.9	270.8	3.0	0.4	0.7
14	20	5	115.9	302.3	0.0	0.3	0.7
15	10	15	103.1	188.1	0.0	0.2	0.8

Table 4.7 below contains the threshold values that were identified based on criterion 2 for each participant. Again, it was determined that there is no consistency in the threshold values among the participants. Also, there were some participants with threshold values that could achieve perfect (= 1) specificity.

Table 4.7 The values of the parameters ($RMSD_{crit}$, ε) selected based on criterion 2 for each participant (study 1)

Partici-pant	$RMSD_{crit}$ (pixel units)	ε (sec)	Duration of False detection (sec)	Duration of Missed detection (sec)	Delay time to detection (sec)	1-Specificity	Sensitivity
1	25	10	26.0	452.1	0.0	≈ 0.0	0.6
2	10	20	0.0	428.2	90.0	0.0	0.6
3				No report of PB			
4	20	15	23.0	530.9	26.0	0.1	0.5
5	15	20	0.0	395.1	163.1	0.0	0.6
6	25	25	32	158.1	87	≈ 0.0	0.9
7	20	15	0.0	582.2	265.1	0.0	0.5
8				No report of PB			
9				No report of PB			
10	25	20	77	422	0.0	0.2	0.6
11	20	15	34	347.1	290.1	0.1	0.7
12	25	15	0	71.1	9.9	0.0	0.9
13	20	10	0	597.6	251.6	0.0	0.5
14	20	15	0.0	542.1	201.9	0.0	0.5
15	15	25	16.1	510.6	8.9	≈ 0.0	0.5

4.5 Discussion

This section discusses the results that have been presented in the previous section. The time to take proper action to keep the target at the center point increased as the task difficulty increased. The target constantly changed its location and the participant needed to continuously monitor its movement and apply appropriate action to bring the target to the center point, whenever there was a need for the correction due to the target's new position. Changing the sensitivity of the joystick added difficulty in controlling the target. Increasing the amount of target movement per update may have increased the time it takes to bring the target within $RMSD_{crit}$. Hence, it became difficult for the participant to promptly apply proper correct adjustment to bring the target back to the center point. Continuation of such failure (delay in correction for each required adjustment) was expected to be captured using the PB detection method. There were some indications that PB can be detected using the PB detection method, particularly when the parameters ($RMSD_{crit}$, ε) of the detection method were calibrated per individual by generating an ROC curve.

The arbitrariness of subjective declaration of his/her PB point may also have negatively affected the effectiveness of the PB detection method, although clear instructions were given to the participants on when to identify PB. Currently, the only available way of knowing when PB occurs is by having the participants declare PB when they subjectively experience it. Hence, such subjective indication was the only available information to evaluate which the combination of the parameters can detect PB effectively. If there is an objective way of knowing when PB actually occurred other than subjective PB declaration, then the parameters of the detection method could be more clearly identified using the detection method.

In addition, it was found that the best combination of $RMSD_{crit}$ and ε values are substantially different for each participant. Although clear instruction was given to the participants that the goal is to keep the target at the center point, participants were performing at different levels. The contributing factors of such individual difference could be due to one’s capability of performing the tasks. In order for the PB detection method to work effectively, participants must show good tracking task performance when they possess the control of the task, whereas some of the participants did not show such performance throughout the whole operation in the study. Establishing thresholds for determining which participants are good candidate for applying the PB detection method would be helpful. One possible approach for determining such thresholds is by setting a fixed $RMSD_{crit}$ value and applying the PB detection method on the participants who can maintain their performance within the set value for the parameter as long as they possess control of the task.

There are also unavoidable false and missed detection of PB in nature using this method. For example, there are occasions where it is hard to distinguish mistakes from PB. The operator may not monitor the task and apply the proper action for the minimum duration of time (ε) , and this may not necessarily be related to PB. For instance, the participant simply had to scratch his/or her arm for the duration of time (ε) when the target deviated beyond $RMSD_{crit}$, then it will be identified as PB, which is a false detection. In addition, the participants were given three tasks to perform concurrently. Participants needed to have good strategy to maintain good performance on all three tasks. However, the participants may not monitor the primary task by not distributing the time to allocate on each task properly and may incidentally spend too much time on the secondary task. There are occasions that PB cannot be detected as well. For example, after the participant brings the target back to the center point, it may take a longer time to deviate beyond the $RMSD_{crit}$, as the target's next movement is determined is based on random chance. Also, the participant may be able to incidentally bring the target back to the center point.

Participants' performance on the secondary tasks was observed. (See Appendix A.) Irrational and random performance in the secondary tasks around PB point was observed for some participants. Moreover, there were differences in how each participant managed the secondary tasks. Some participants showed poor performance in either resource management task, the system monitoring task, or both. Furthermore, there were participants who did not show poor performance in the secondary tasks at all. (See Appendix A.) It is observed that participants 2, 4, 5, 8, 10, and 12 dropped the resource management task at or around the subjectively declared PB point. It is observed that participants 1, 6, 7, 10, 11, 14, and 15 opportunistically responded to the stimuli in the system monitoring task throughout the operation. The strategic/or tactical strategies to maintain adequate performance on the tracking task could have affected the effectiveness of the PB detection method on detecting PB based on the primary task performance. However, the difficulty of using an indication of the changes in the strategy for performing the secondary task as an indication/or predictor of PB point was identified. There was no consistency in how the participants were changing the strategies and when they would change it.

System designers can build an effectively operational system as long as the specificity or sensitivity of the detection method are consistent over time. One example is a warning system that triggers a warning when PB is detected. In the operation of such a warning system, PB may get falsely detected and a warning may be triggered at a non-PB point. Even if the warning has been falsely triggered, the human operator can have awareness of his/her low level of performance.

CHAPTER 5. STUDY TWO

5.1 Overview

Study 2 was conducted: 1) to repeat the demonstration of inducing PB, 2) to test whether the threshold parameters established from study 1 for the PB detection method can be used in a subsequent study, or whether they have to be re-calibrated for each study, and 3) to examine whether a specific physiological measure (pulse rate) can be used to identify the subjectively-reported PB point. The three tasks were presented to the participants in an identical way to the first study. Both the performance and the physiological data were collected. The performance data was investigated to evaluate whether the threshold values, which were identified in study 1, could be used again in study 2. In this study, a physiological measure was collected to search for an additional indication of PB.

5.2 Method

5.2.1 Participants

The same 15 (13 male + 2 female) participants from the first study were asked to participate in the second study on the same day as their participation in the first study.

5.2.2 Tools (Physiological Measures) and Procedures

Participants' pulse rates were gathered using the BioHarness™ 3 from Zephyr's BioHarness technology (Figure 5.1). This device was placed on the participant's chest using a strap, which incorporated electrocardiography (ECG) and breathing detection sensors. The collected data were analyzed through OmniSense Analysis software from the same company.

Figure 5.1 The BioHarness™ 3

5.2.3 Independent Variables

The participants were asked to perform the same three tasks from the MATB-II as study 1. The tasks were operated in an identical way to study 1: the difficulty of the primary task increased every two minutes while the difficulty of the secondary tasks remained static.

5.2.4 Dependent Variables

There were four dependent variables: 1) time of PB that the participant verbally indicated, 2) RMSE the tracking task (pixel unit), 3) errors in the secondary tasks (resource management task and system monitoring task), and 4) pulse rate (bpm).

5.3 Hypotheses

First, the demonstration of inducing PB was repeated in the same environment as Study 1.

Hypothesis 2a: Increasing workload will induce subjectively-identified PB.

Next, evaluation of the PB detection method on the tracking task performance was conducted to identify whether the threshold parameters established from study 1 for the PB detection method can be used in a subsequent study, or whether they should be re-calibrated for each study.

Hypothesis 2b: The threshold parameters ($RMSD_{crit}$ and ε) established from study 1 for the PB detection method can be used in a subsequent study.

Second, pulse rate was introduced as an additional predictor of PB. It was examined whether the same approach for objectively identifying PB on performance data can also be used on the pulse rate data.

The assumption of taking the same approach for objectively identifying PB on the pulse rate data was that there is a direct mapping between pulse rate and performance. Hence, Equation 4.3 has been modified as follows:

$$(Pulse\ rate > Pulse\ rate_{crit}) \cap (\Delta t_{pulse\ rate} \geq \varepsilon) \quad (5.1)$$

In Equation 5.1 above, PB occurs when pulse rate exceeds a maximum pulse rate threshold ($Pulse\ rate_{crit}$) that directly maps to the minimum performance level that an operator is required to maintain. $\Delta t_{pulse\ rate}$ is a duration of time when the current pulse rate is greater than $Pulse\ rate_{crit}$. ε is a minimum duration of time allowance for the pulse rate to go below the $Pulse\ rate_{crit}$.

> *Hypothesis 2c: There exist criteria ($Pulse\ rate_{crit}$ and ε) such that the point corresponding to $(P < P_{crit}) \cap (\Delta t \geq \varepsilon)$ matches the subjectively-identified performance breakdown point.*

After examining whether the same approach for objectively identifying PB on performance data can also be used on the pulse rate data, an additional exploratory test was conducted. The relationship between pulse rate and tracking task performance was characterized by performing a K-means clustering analysis.

5.4 Results

5.4.1 Overview

This section presents the results of the data analysis for Study 2. The results of the hypothesis testing were: 1) increasing workload can induce subjectively-identified PB, although it may not be generalizable, 2) the threshold parameters established in study 1 for the PB detection method cannot be used in a subsequent study, and suggest that it may require re-calibration for each study, and 3) there does not appear to be any evidence that pulse rate data can be used to detect PB.

5.4.2 Hypothesis 2a

The same 12 out of 15 participants declared PB as in study 1, which supports the hypothesis 2a (See Table 5.1).

Table 5.1 Summary statistics of participant's tracking task performance (pixel units) and PB (PB = Yes or Non-PB = No)

Participant	Mean	SD	Median	PB
1	26.1	18.9	22.0	Yes
2	17.8	10.5	15.6	Yes
3	35.1	28.9	26.5	No
4	21.5	12.3	19.2	Yes
5	33.1	22.7	27.0	Yes
6	27.6	17.5	24.1	Yes
7	29.5	19.4	25.2	Yes
8	19.8	11.8	17.7	No
9	19.0	10.9	16.7	No
10	34.8	22.7	29.8	Yes
11	21.6	14.6	17.9	Yes
12	29.0	17.1	25.1	Yes
13	25.0	18.7	20.7	Yes
14	21.5	12.9	19.3	Yes
15	19.5	10.7	17.7	Yes

5.4.3 Hypothesis 2b

The sensitivity and the specificity for each participant were computed. A ROC curve was constructed for each participant to investigate how various threshold values affect PB detection. (See Appendix B.)

Table 5.2 presents the threshold values that are identified based on criterion 1 for study 1 and study 2. The threshold parameters established from study 2 are different from the parameters identified in study 1.

Table 5.2 The values of the parameters ($RMSD_{crit}, \varepsilon$) selected based on criterion 1 for each participant (study 1 and 2)

Partici-pant	Study 1				Study 2			
	$RMSD_{crit}$ (pixel units)	ε (sec)	1-Spec.	Sens.	$RMSD_{crit}$ (pixel units)	ε (sec)	1-Spec.	Sens.
1	15.0	10.0	0.3	0.8	10.0	10.0	0.4	0.8
2	10.0	10.0	0.2	0.8	10.0	10.0	0.0	0.4
3	No report of PB							
4	10.0	25.0	0.3	0.6	10.0	25.0	0.1	0.7
5	15.0	10.0	0.1	0.8	15.0	20.0	0.2	0.8
6	25.0	25.0	≈ 0.0	0.9	15.0	25.0	0.2	0.9
7	15.0	10.0	0.2	0.7	15.0	25.0	0.0	0.8
8	No report of PB							
9	No report of PB							
10	25.0	10.0	0.3	0.7	25.0	25.0	≈ 0.0	0.8
11	10.0	25.0	0.2	0.9	10.0	25.0	0.0	0.7
12	20.0	15.0	≈ 0.0	1.0	20.0	25.0	0.0	0.6
13	10.0	15.0	0.4	0.7	20.0	15.0	0.0	0.3
14	20.0	5.0	0.3	0.7	15.0	20.0	0.0	0.4
15	10.0	15.0	0.2	0.8	15.0	10.0	0.1	0.7

Table 5.3 presents the threshold values that are identified based on criterion 2 for study 1 and study 2. Again, the threshold parameters established from study 2 are

different from the parameters identified in study 1, which suggests that threshold parameters established in study 1 for the PB detection method cannot be used in a subsequent study.

Table 5.3 The values of the parameters ($RMSD_{crit}, \varepsilon$) selected based on criterion 2 for each participant (Study 1 and 2)

Partici-pant	Study 1				Study 2			
	$RMSD_{crit}$ (pixel units)	ε (sec)	1-Spec.	Sens.	$RMSD_{crit}$ (pixel units)	ε (sec)	1-Spec.	Sens.
1	25.0	10.0	≈ 0.0	0.6	15.0	25.0	0.1	0.7
2	10.0	20.0	0.0	0.6	20.0	25.0	0.0	0.8
3	No report of PB							
4	20.0	15.0	0.1	0.5	15.0	25.0	0.0	0.5
5	15.0	20.0	0.0	0.6	15.0	25.0	0.1	0.7
6	25.0	25.0	≈ 0.0	0.9	20.0	25.0	0.0	0.8
7	20.0	15.0	0.0	0.5	25.0	15.0	0.0	0.8
8	No report of PB							
9	No report of PB							
10	25.0	20.0	0.2	0.6	25.0	25.0	≈ 0.0	0.8
11	20.0	15.0	0.1	0.7	10.0	25.0	0.0	0.7
12	25.0	15.0	0.0	0.9	20.0	25.0	0.0	0.6
13	20.0	10.0	0.0	0.5	20.0	15.0	0.0	0.3
14	20.0	15.0	0.0	0.5	15.0	20.0	0.0	0.4
15	15.0	25.0	≈ 0.0	0.5	20.0	10.0	0.0	0.6

The threshold values that were identified based on criterion 1 from the first study were applied to the data collected during the second study. Table 5.4 provides a summary of the analysis.

Table 5.4 The performance of re-using the threshold values selected (criterion 1) from study 1 on the data collected in study 2

Partici-pant	$RMSD_{crit}$ (pixel units)	ε (sec)	Total duration of false detection (sec)	Total duration of missed detection (sec)	Delay time to detection (sec)	1-Spec.	Sens.
1	15.0	10.0	27.0	632.0	66.5	0.1	0.4
2	10.0	10.0	45.4	409.0	80.1	0.1	0.5
3				No report of PB			
4	10.0	25.0	134.0	150.0	0.0	0.3	0.7
5	15.0	10.0	193.0	225.0	0.0	0.5	0.7
6	25.0	25.0	38.9	268.0	N\A	0.1	0
7	15.0	10.0	307.0	39.0	0.0	0.5	0.9
8				No report of PB			
9				No report of PB			
10	25.0	10.0	291.0	122.0	0.0	0.3	0.5
11	10.0	25.0	22.0	282.0	0.0	0.1	0.6
12	20.0	15.0	68.0	300.0	10.0	0.1	0.3
13	10.0	15.0	73.0	208.0	97.0	0.4	0.8
14	20.0	5.0	31.0	549.0	166.0	0.1	0.3
15	10.0	15.0	158.0	219.0	47.0	0.3	0.6

The threshold values that were identified based on criterion 2 from the first study were applied to the data collected during the second study. Table 5.5 provides a summary of the analysis.

Table 5.5 The performance of re-using the threshold values selected (criterion 2) from study 1 on the data collected in study 2

Parti-cipant	$RMSD_{crit}$ (pixel units)	ε (sec)	Total duration of false detection (sec)	Total duration of missed detection (sec)	Delay time to detection (sec)	1-Spec.	Sens.
1	25.0	10.0	0.0	776.0	78.5	0.0	0.1
2	10.0	20.0	0.0	509.0	331.0	0.0	0.4
3				No report of PB			
4	20.0	15.0	35.0	497.0	N\A	0.1	0.0
5	15.0	20.0	118.0	249.0	0.0	0.3	0.6
6	25.0	25.0	38.9	0.0	N\A	0.1	0.0
7	20.0	15.0	71.0	367.0	21	0.1	0.4
8				No report of PB			
9				No report of PB			
10	25.0	20.0	75.0	843.0	0.0	≈ 0.0	≈ 0.0
11	20.0	15.0	0.0	0.0	0.0	0.0	0.0
12	25.0	15.0	20.0	416.0	83.0	≈ 0.0	0.1
13	20.0	10.0	14.0	631.0	282.9	0.1	0.3
14	20.0	15.0	0.0	0.0	0.0	0.0	0.0
15	15.0	25.0	0.0	0.0	0.0	0.0	0.0

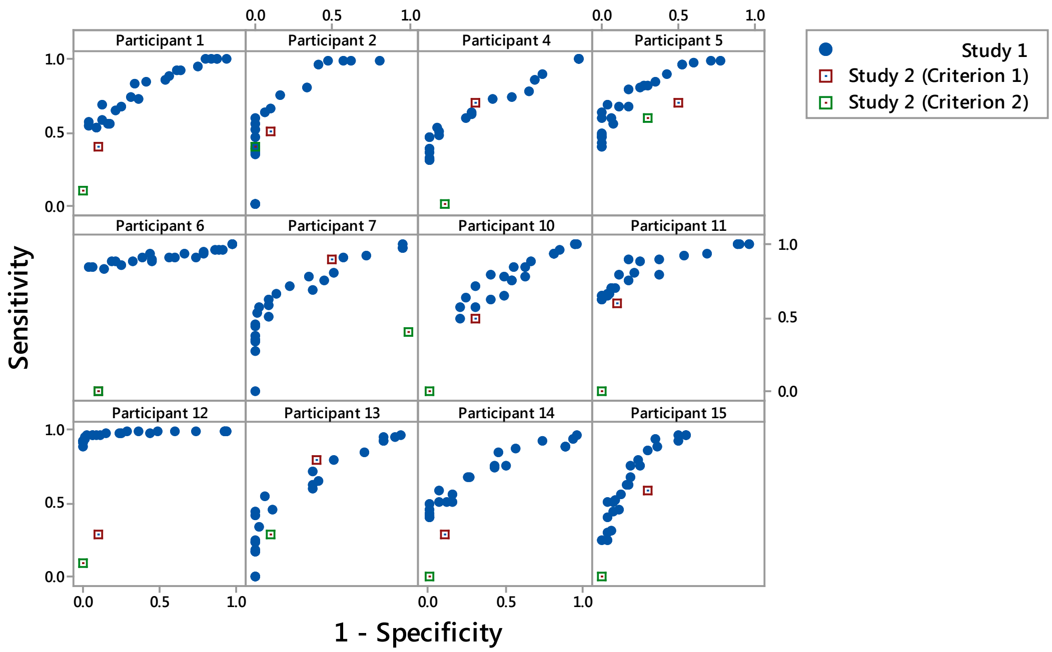


Figure 5.2 Receiver operating characteristic (ROC) curve: evaluation of parameters ($RMSD_{crit}$, ε) on PB detection.

Figure 5.2 shows how re-using the threshold parameters identified from study 1 for all participants performed. In general, re-using the parameters is found to be ineffective. For some participants, there was inconsistency in performance between studies (study 1 and 2), which could have been the problem with re-using the established parameters for those participants.

For some participants, there was inconsistency in performance between studies (study 1 and 2), which could have been the problem with re-using the established parameters for those participants.

The histogram of the tracking task (RMSD) was constructed to visually observe for any abnormality in the data.

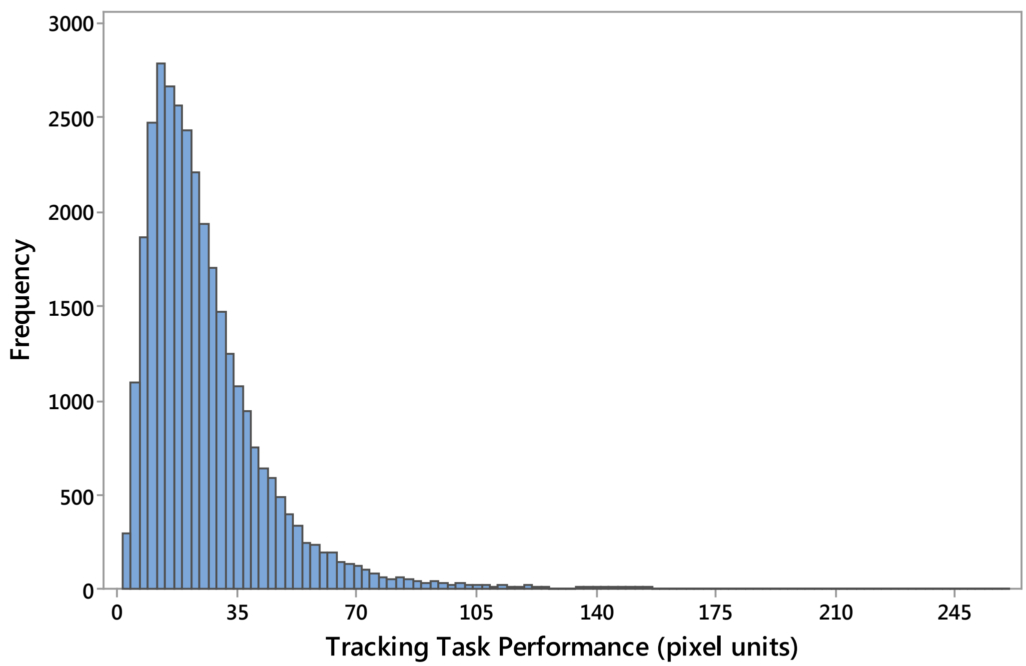


Figure 5.3 Histogram of the tracking task performance (study 1 and 2)

In Figure 5.3, it can be observed that the distribution of the tracking task performance has a left skew with a long right tail. The results of the normality test (the Anderson-Darling test) indicate that the tracking task performance data is not normally distributed. (See Figure 5.3).

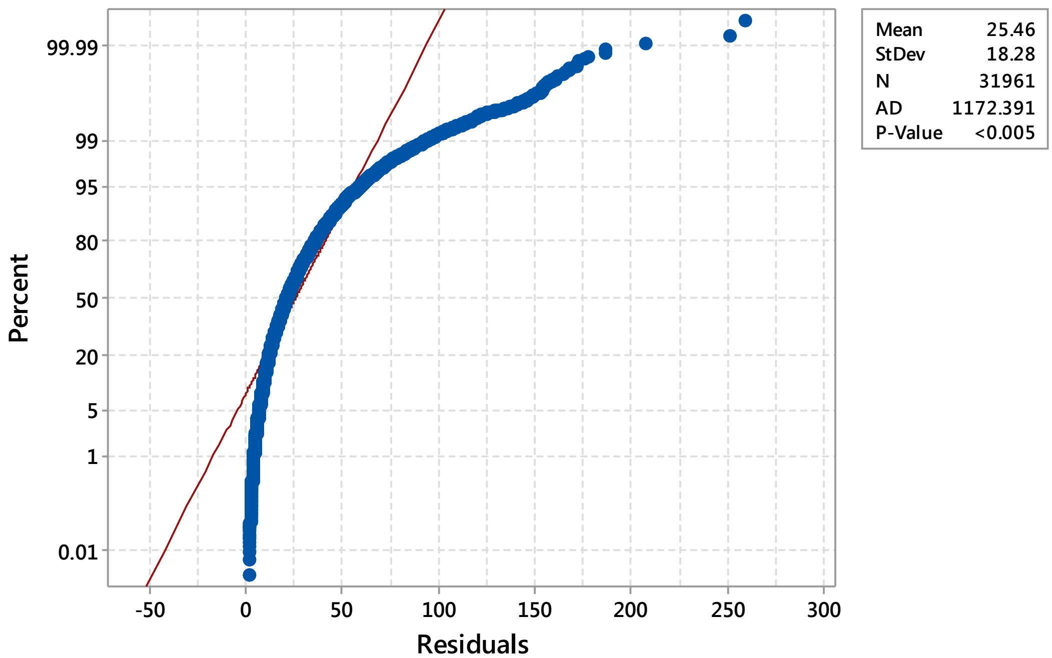


Figure 5.4 Anderson-Darling test on the tracking task performance (study 1 and 2)

Levene's test indicates there is constant variance between study 1 and 2 ($p < 0.005$).

Therefore, the median values (Table 5.6) from the data were used as measures of central tendency. The results of a non-parametric test (Mann-Whitney) show that participants performed better during study 2 compared to study 1 (W = 257352970.5, p – value = 0.0238).

Table 5.6 Summary statistics for primary task performance (pixel units) (study 1 vs. study 2)

	Mean	SD	Median
Study 1	25.6	18.2	21.2
Study 2	25.3	18.3	20.9

Figure 5.5 also shows that there is a relatively small improvement in performance between study 1 and study 2.

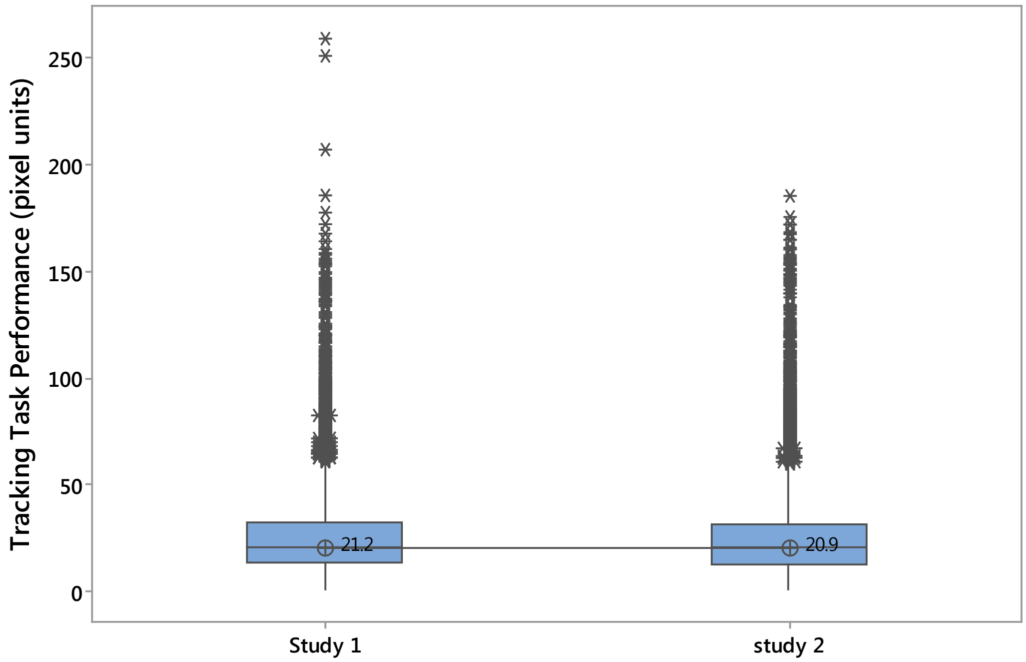


Figure 5.5 Tracking task performance (RMSD): study 1 vs. study 2

5.4.4 Hypothesis 2c

Next, analysis was conducted on each individual's pulse rate data. Figure 5.6 is the box-plot of the average pulse rate under the different task difficulty level.

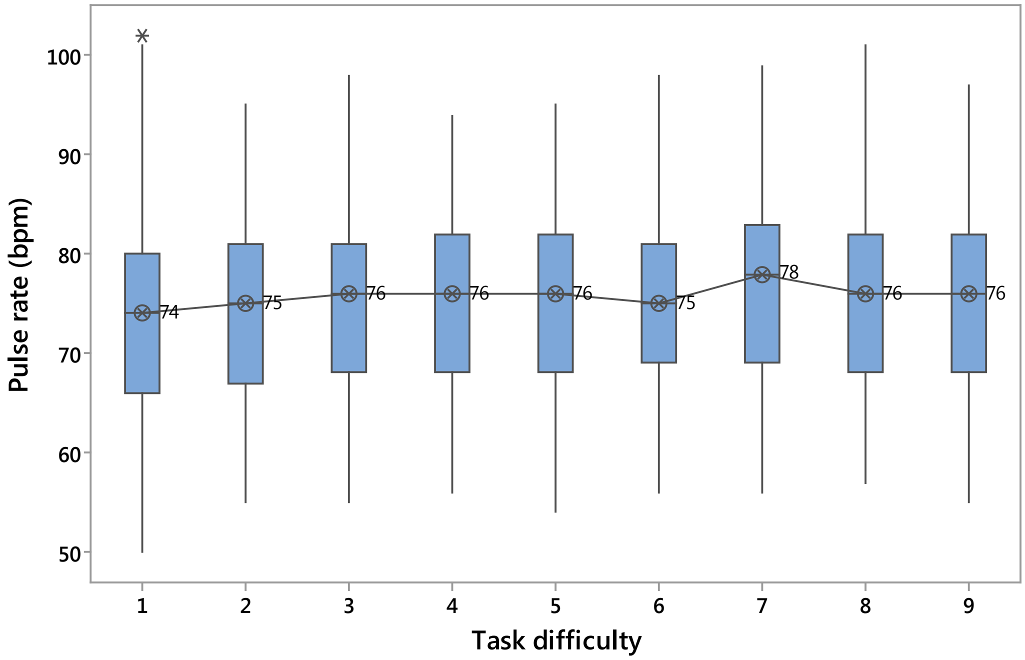


Figure 5.6 Average pulse rate vs. task difficulties

The histogram of pulse rate (Figure 5.7) was constructed to visually examine whether the data are skewed or any outliers exist in the data. The data reveals that there are multiple peaks in the histogram, which might be due to significant individual differences.

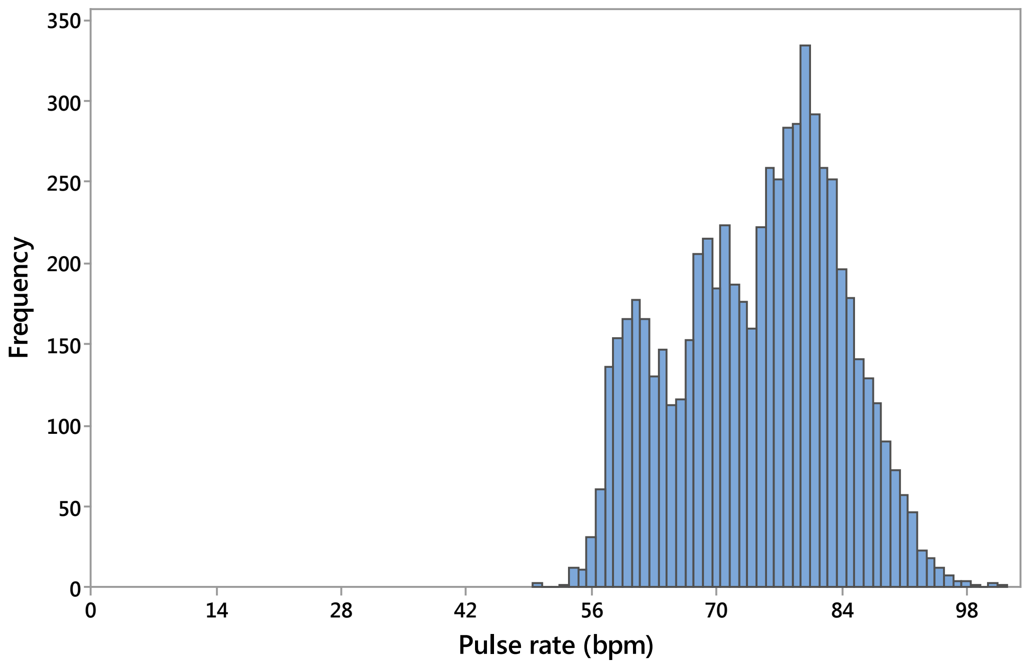


Figure 5.7 Histogram of pulse rate data

Table 5.7 Pulse rate (bpm) summary statistics (Pre-PB vs. Post-PB)

	Mean	SD	Median
Pre-PB	71.7	9.0	71.0
Post-PB	73.7	8.3	75.0

Each participant's pulse rate data was investigated separately. The histogram of pulse rate (Figure 5.7) indicates that there are large individual differences. Table 5.8 represents the summary statistics of each participant's pulse rate data. In the table, it can be determined that there are large differences in each individual's pulse rate data. Participant 7's average pulse rate is 60.2 bpm, while participant 8's average pulse rate is 86.5 bpm. The difference is 26.3 bpm.

Table 5.8 Summary statistics for participant's pulse rate (bpm) and PB (PB = Yes, Non – PB = No)

Participant	Mean	SD	Median	Resting Pulse Rate	PB
1	78.3	4.0	69.0	75.7	Yes
2	70.6	4.5	70.0	63.0	Yes
3	74.1	5.1	73.0	66.2	No
4	60.7	3.3	60.0	60.0	Yes
5	70.9	5.8	71.0	67.6	Yes
6	63.4	4.1	63.0	55.8	Yes
7	60.2	2.4	60.0	56.9	Yes
8	86.5	3.4	87.0	74.0	No
9	85.5	5.1	85.0	71.7	No
10	82.8	4.1	82.0	71.5	Yes
11	78.9	3.6	79.0	75.4	Yes
12	82.0	3.9	82.0	77.0	Yes
13	75.9	4.7	76.0	71.8	Yes
14	69.5	2.7	69.0	66.9	Yes
15	79.6	3.2	79.0	73.0	Yes

Hence, each participant's pulse rate changes during the run were plotted individually.

(Figure 5.8.)

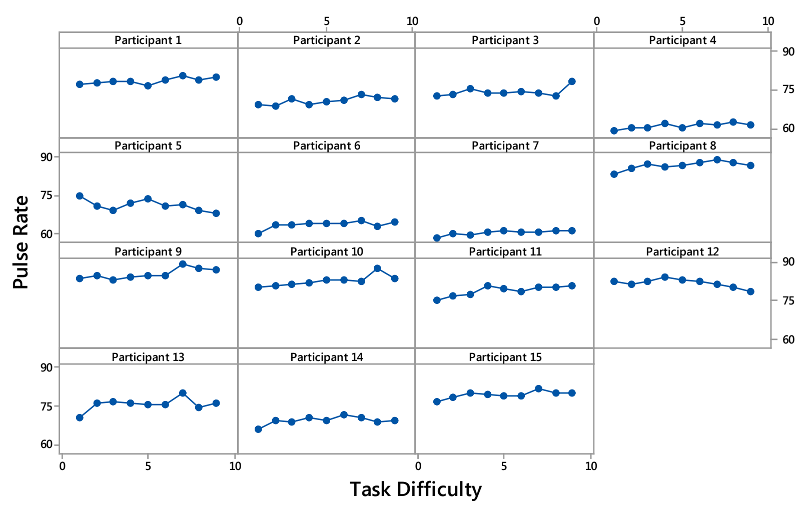


Figure 5.8 Pulse rate (bpm) vs. task difficulty

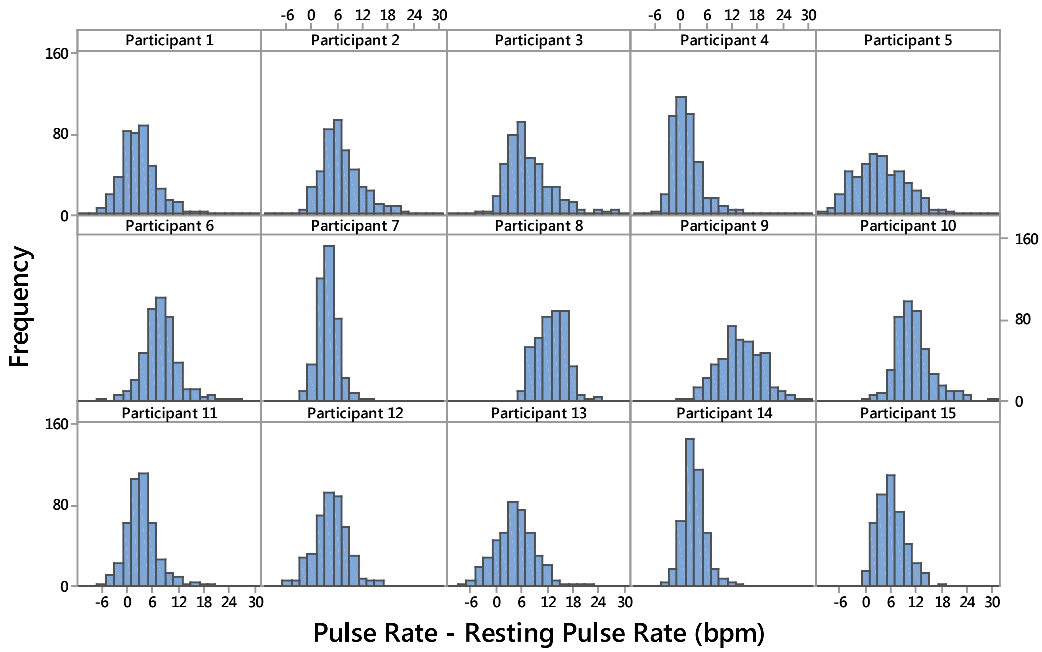


Figure 5.9 Pulse rate – resting pulse rate (bpm)

Figure 5.9 indicates the histogram of each individual's pulse rate data that has been normalized by subtracting the resting pulse rate from pulse rate. Next, the correlation between the average of the values (pulse rate – resting pulse rate) and the task difficulty was tested, where the result indicates that there is a high correlation (Pearson correlation coefficient = 0.8, p – value = 0.01).

The correlation between each participant's pulse rate and task difficulty was looked at individually this time. Table 5.9 summarizes the results of the correlation test.

Table 5.9 Results of the correlation test (pulse rate vs. task difficulty)

Participant	Pearson Correlation Coefficient	p-value
1	0.7	0.0
2	0.8	0.0
3	0.5	0.2
4	0.7	0.0
5	-0.6	0.1
6	0.6	0.1
7	0.9	0.0
8	0.7	0.0
9	0.8	0.2
10	0.8	0.0
11	0.8	0.0
12	-0.7	0.0
13	0.4	0.3
14	0.5	0.2
15	0.6	0.1

The pulse rate data of participants 3, 5, 6, 13, 14, and 15 indicate that their pulse rates do not correlate with the increase in task difficulty. The pulse rate data of participants 5 and 12 decrease toward the end, although the participants continued to perform the task. Hence, the Pearson correlation coefficient value came out to be negative.

Each participant's pulse rate data prior to PB and after PB were further analyzed and presented in Table 5.10. The data shows that pulse rate increased after the PB point, but only by a small increment (2 bpm in average).

Table 5.10 Pulse rate (bpm): pre-PB vs. post-PB

Participant	Mean	SD	Median	Pulse rate – Resting Pulse rate	% increase relative to the resting pulse rate	Difference between Pre-PB and Post-PB pulse rate
1 (Pre-PB)	76.4	3.2	76.5	0.7	0.9	2.3
1 (Post-PB)	78.7	4.1	78.0	3	4.0	
2 (Pre-PB)	69.3	4.2	69.0	6.3	10.0	1.7
2 (Post-PB)	71.0	4.5	70.0	8	12.7	
4 (Pre-PB)	60.1	3.1	59.0	0.1	0.2	1.1
4 (Post-PB)	61.2	3.4	61.0	1.2	2.0	
5 (Pre-PB)	71.0	5.5	71.0	3.4	5.0	-0.2
5 (Post-PB)	70.8	5.9	70.0	3.2	4.7	
6 (Pre-PB)	63.5	4.1	64.0	7.7	13.8	-0.1
6 (Post-PB)	63.4	4.4	63.0	7.6	13.6	
7 (Pre-PB)	59.8	2.6	60.0	2.9	5.1	1
7 (Post-PB)	60.8	2.0	61.0	3.9	6.9	
10 (Pre-PB)	81.9	3.5	82.0	10.4	14.6	4.1
10 (Post-PB)	86.0	4.6	85.0	14.5	20.3	
11 (Pre-PB)	76.9	3.0	77.0	1.5	2.0	3.1
11 (Post-PB)	80.0	3.4	80.0	4.6	6.1	
12 (Pre-PB)	83.0	3.1	83.0	6	7.8	-2.4
12 (Post-PB)	80.6	4.5	80.0	3.6	4.7	
13 (Pre-PB)	72.6	5.7	71.0	0.8	1.1	4.1
13 (Post-PB)	76.7	4.1	76.0	4.9	6.8	
14 (Pre-PB)	68.1	2.6	69.0	1.2	1.8	1.9
14 (Post-PB)	70.0	2.5	69.0	3.1	4.6	
15 (Pre-PB)	78.9	3.0	79.0	5.9	8.1	1.5
15 (Post-PB)	80.4	3.3	80.0	7.4	10.1	

The same approach for objectively identifying PB on performance data was applied to determine whether it can also be used on the pulse rate data. The value of $Pulse\ rate_{crit}$was set at the center point between the mean pulse rates of Pre-PB and Post-PB. The different ε values (5, 10, 15, 20, and 25) were applied to see which value captures the changes the best. Table 5.11 reports the results of testing the PB detection method on the pulse rate data.

Table 5.11 The best values of the parameters ($Pulse\ rate_{crit}$, ε)

Partici-pant	$Pulse\ rate_{crit}$ (bpm)		ε (seconds)				
			5	10	15	20	25
1	77.6	1-Specificity	0.0	0.0	0.0	0.0	0.0
		Sensitivity	0.5	0.4	0.3	0.2	0.1
2	70.2	1-Specificity	0.4	0.0	0.0	0.0	0.0
		Sensitivity	0.3	0.2	0.1	0.1	0.1
4	60.7	1-Specificity	0.1	0.1	0.0	0.0	0.0
		Sensitivity	0.4	0.3	0.3	0.2	0.1
5	70.9	1-Specificity	0.5	0.4	0.3	0.2	0.2
		Sensitivity	0.5	0.3	0.3	0.1	0.0
6	63.5	1-Specificity	0.5	0.3	0.2	0.1	0.0
		Sensitivity	0.5	0.4	0.3	0.2	0.1
7	60.3	1-Specificity	0.3	0.2	0.1	0.1	0.0
		Sensitivity	0.5	0.2	0.1	0.1	0.1
10	84.0	1-Specificity	0.2	0.1	0.0	0.0	0.0
		Sensitivity	0.3	0.3	0.1	0.1	0.1
11	78.5	1-Specificity	0.2	0.2	0.1	0.1	0.1
		Sensitivity	0.6	0.5	0.4	0.3	0.3
12	81.8	1-Specificity	0.6	0.5	0.4	0.3	0.3
		Sensitivity	0.1	0.1	0.1	0.1	0.1
13	74.7	1-Specificity	0.3	0.2	0.2	0.2	0.2
		Sensitivity	0.1	0.0	0.0	0.0	0.0
14	69.1	1-Specificity	0.2	0.1	0.1	0.1	0.1
		Sensitivity	0.2	0.1	0.1	0.1	0.1
15	79.7	1-Specificity	0.4	0.2	0.2	0.1	0.1
		Sensitivity	0.0	0.0	0.0	0.0	0.0

Figure 5.10 shows the results in Table 5.11. The ROC curves for each participant are presented to visually show the performance of the different threshold values on detecting PB. The value in the top right corner indicates the different ε values, which are color-coded.

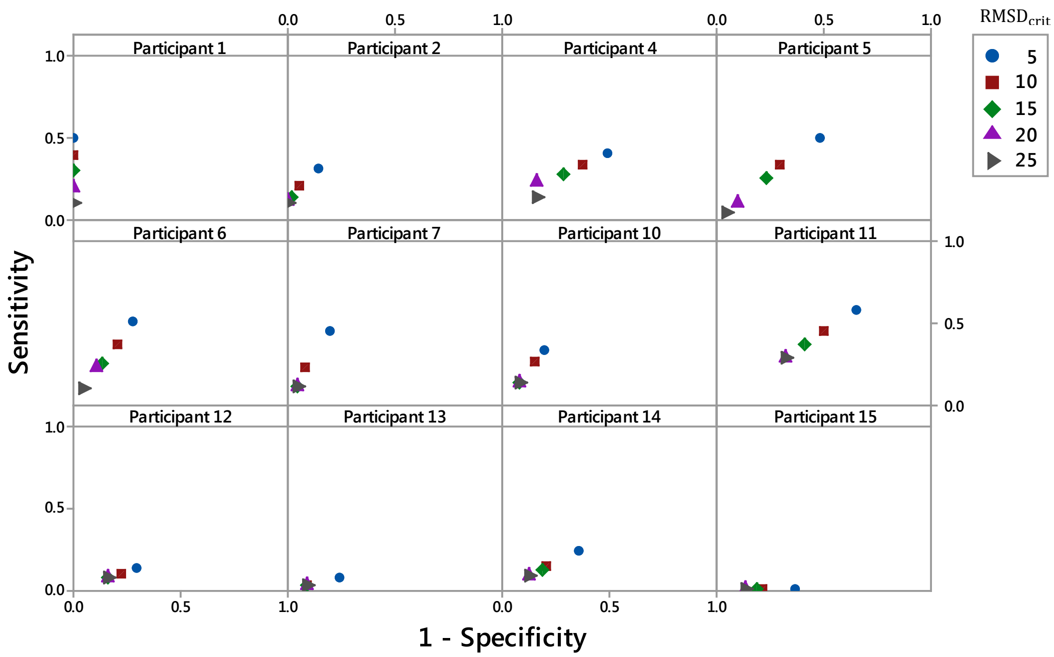


Figure 5.10 ROC curve: evaluation of parameters ($Pulse\ rate_{crit}$, ε)

The ROC curve in Figure 5.10 seems to provide no clear evidence that pulse rate can be used to detect PB, since the points lie along the diagonal, which represents simple chance detection.

After examining whether same approach for objectively identifying PB on performance data can also be used on the pulse rate data, the relationship between pulse rate and tracking task performance was characterized by performing K-means clustering analysis.

The clustering method was used on participants' pulse rate data. Figure 5.11 shows the scatter plot of pulse rate (bpm) vs. tracking task (RMSD). Each dot represents the average value of the data collected for each 5 second interval. The plots are categorized into two clusters (PB vs. Non-PB), and the results indicate no clear relationship between changes in the tracking task performance and the pulse rate.

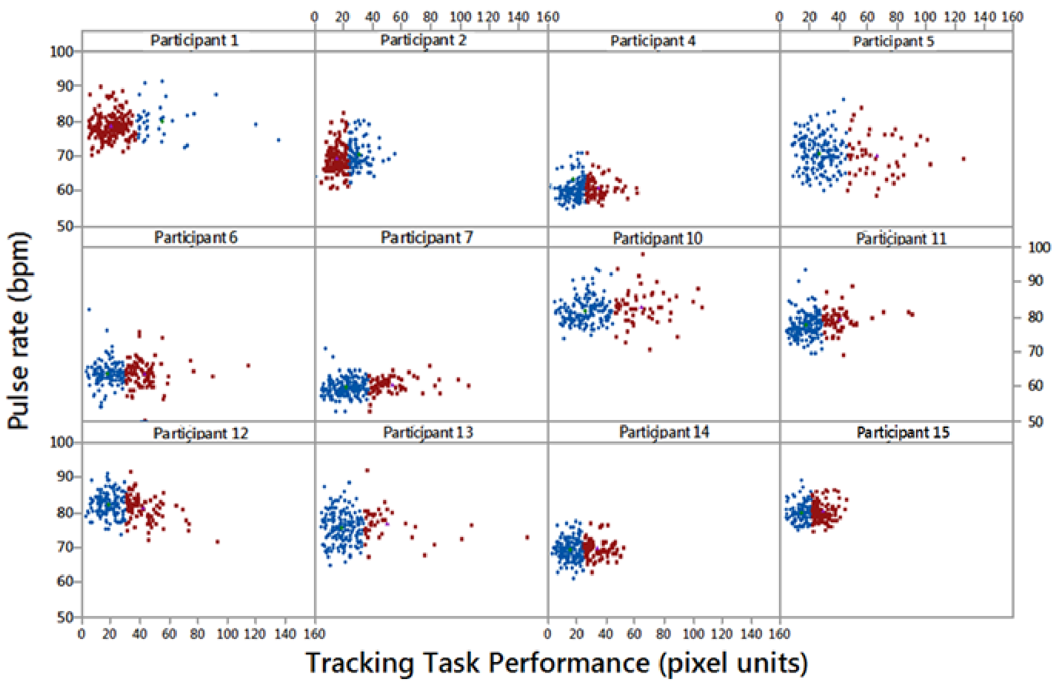


Figure 5.11 Pulse rate (bpm) vs. tracking task performance (pixel units)

There appears to be no relationship between pulse rate and tracking task performance on the basis of this experiment (See Figure 5.11).

5.5 Discussion

It was demonstrated again that PB can be induced by increasing workload as the same 12 out of 15 participants reported being in a state of PB. The threshold parameters established ($RMSD_{crit}$, ε) in study 1 for the PB detection method cannot be used in a subsequent study, and suggest that it may require re-calibration for each study. In study 1, the best combination of $RMSD_{crit}$ and ε values were identified to be different for each participant. Although there were clear instructions given to the participants that the goal is to keep the target at the center point, participants were performing at different levels, which would have negatively affected the identification of the effective combination of the parameters in detecting PB. There was also inconsistency in how some of the participants performed between studies. This indicates performance characteristics of PB may not be consistent over time.

In the second part of the study, a correlation between pulse rate and task difficulty was first identified. This supports the conclusion that pulse rate is a valid metric for measuring task difficulty changes that have been suggested by previous research (Jorna, 1993; Roscoe, 1993; Light & Obrist, 1983; Wright, Contrada, & Patane, 1986; Backs et al., 2003; Chen et al., 2008; Haarman et al., 2009; Ting et al., 2010; Lagu, Landry, & Yoo; 2013). However, there were cases where pulse rate and task difficulty was found to be negatively correlated (participant 5 and 12). Although sufficient resting time was provided between the practice run and the actual data collection run, such results might be due to nervousness and agitation in the beginning of the run.

The average difference in each individual's pulse rate data after PB and before PB was found to be approximately 2.0 beats per minute (bpm) and ranged from a 2.0 % to 20.3 % increase depending on one's resting pulse rate. Such small increments in pulse rate after PB suggests that it might not be adequate to detect such differences in real time, as the effect of sinus arrhythmia (i.e., the normal increase in pulse rate occurring during inspiration) is about 15 % (Lagu, Landry & Yoo).

The evaluation of pulse rate as an objective way of detecting PB using the same PB detection method appeared to show no evidence that pulse rate could be used. The ROC curves generated for each participant's data show that the performance of detecting PB using the same approach was no better than random chance when the method was applied on the pulse rate data. Furthermore, there appears to be no relationship between pulse rate and tracking task performance on the basis of this experiment.

CHAPTER 6. STUDY THREE

6.1 Overview

In the third study, the same 15 participants from the first two studies were asked to perform only the primary task, where the task difficulty increased in a same manner as the previous studies. The third study was conducted: 1) to determine if PB is induced by the primary task workload or it is affected by the presence of the secondary tasks, and 2) to re-test whether the threshold parameters established from study 1 can be used in a subsequent study. The impact of the secondary task on PB was investigated.

6.2 Method

6.2.1 Participants

The same 15 participants were asked to participate in the third study on the same day they partook in the first and second studies. Before data collection began, the participants were provided with the purpose of the third study and informed that they were free to withdraw at any time.

6.2.2 Tools and Procedures

The third study was conducted in an identical way to the second study, except that the participants were provided with only the tracking task from the MATB-II. All procedures remained the same as the second study.

6.2.3 Independent Variables

The task difficulty of the primary task (tracking task) increased every two minutes. No secondary tasks were provided.

6.2.4 Dependent Variables

There were two dependent variables: 1) time of PB that the participant verbally indicated, and 2) root mean square error (RMSE) of the tracking task (pixel units).

6.3 Hypotheses

First, it was determined whether the secondary task has an impact on PB.

Hypothesis 3a: PB will occur without the secondary task being present.

Then, further testing was done to determine whether the threshold parameters established from study 1 for PB detection could be used in a subsequent study. If the participants did not subjectively declare PB, then the established parameters should not have detected PB in the data.

Hypothesis 3b: The threshold parameters ($RMSD_{crit}$ and ε) established from study 1 for the PB detection method can be used in a subsequent study (Study 3).

6.4 Results

6.4.1 Overview

This section presents the results of the data analysis for study 3. The results of the hypothesis testing are: 1) PB did not occur without the secondary tasks present, even though the primary task workload was the same, and 2) the threshold parameters established from study 1 for the PB detection method cannot be used in a subsequent study.

6.4.2 Hypothesis 3a

The first part of analysis was centrally focusing on identifying whether there is an improvement in the tracking task performance in the third study compared to the performance in the second study.

First, it was determined that none of the participants in the third study experienced PB, which supports the conclusion that there is a significant impact of the secondary task on PB.

A histogram of the tracking performance (RMSD) is constructed in Figure 6.1, where it is observed that the distribution of the tracking task performance has a left skew with a long right tail.

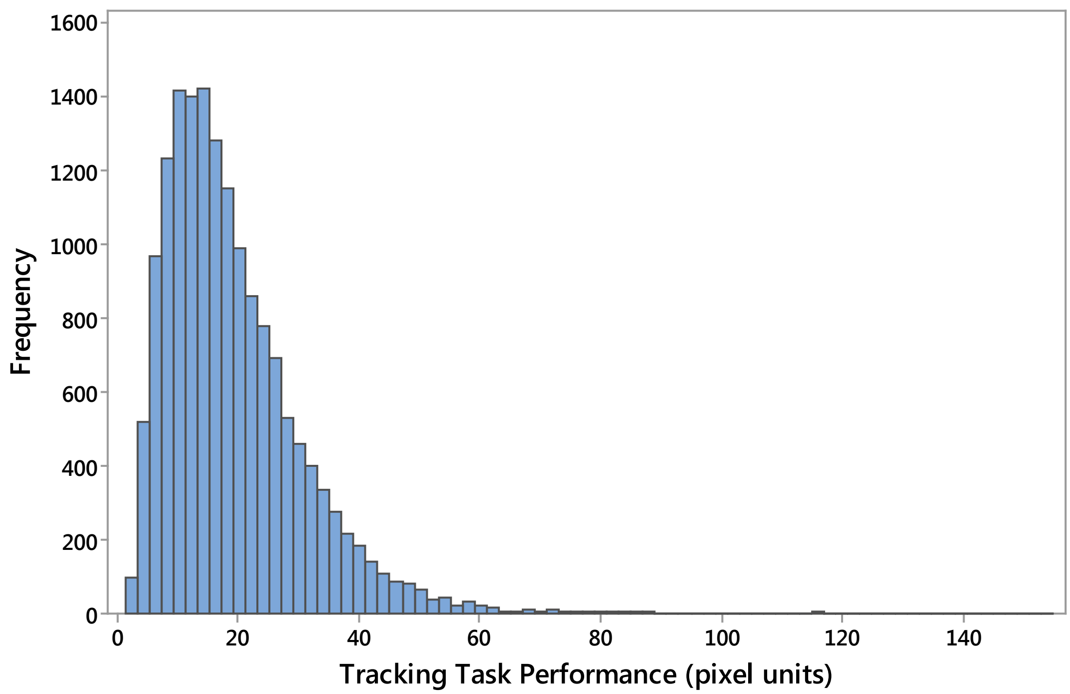


Figure 6.1 Histogram of tracking task performance

An Anderson-Darling test indicated that tracking task performance is not normally distributed ($p < 0.03$).

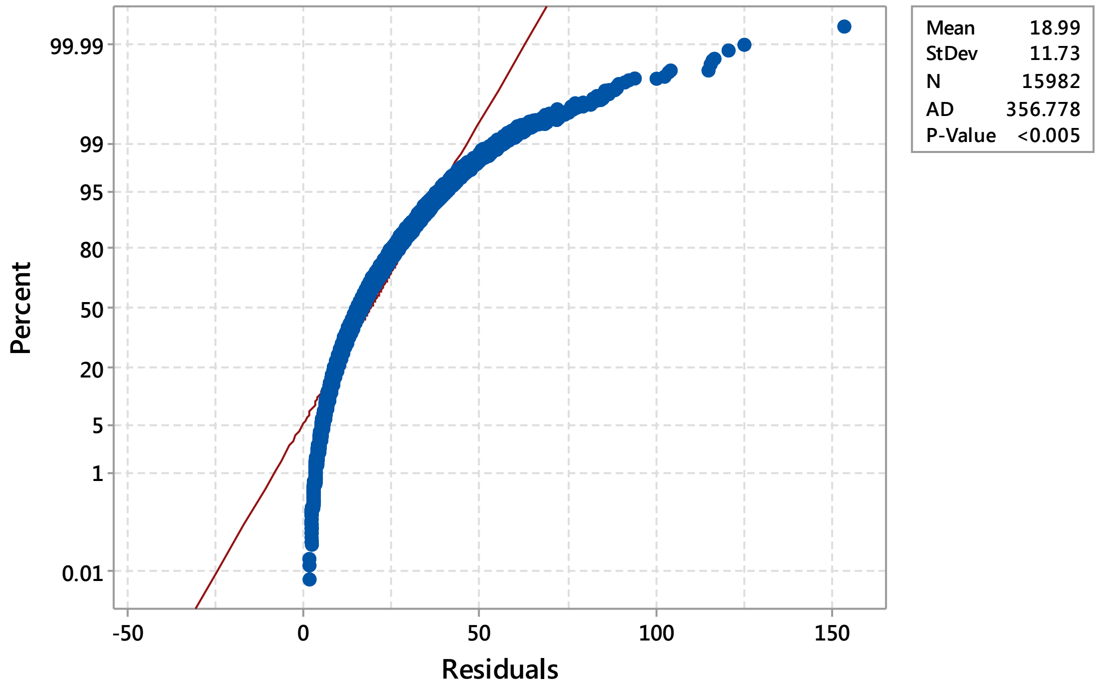


Figure 6.2 Anderson-Darling test on tracking task performance (pixel units)

The Levene's test was performed to examine the equal variances of the tracking task performance data from study 2 and study 3 and it is determined that the data lacks in homoscedasticity.

Therefore, the median values (Table 6.1) from the data were used as measures of central tendency. The results of a non-parametric test (Mann-Whitney) show that participants performed significantly better during study 3 compared to study 2 (W = 288243293.5, p = 0.0000).

Table 6.1 Summary Statistics for Primary Task Performance (pixel units) (Study 2 vs. Study 3)

	Mean	SD	Median
Study 2	25.3	18.4	20.9
Study 3	19.0	11.7	16.4

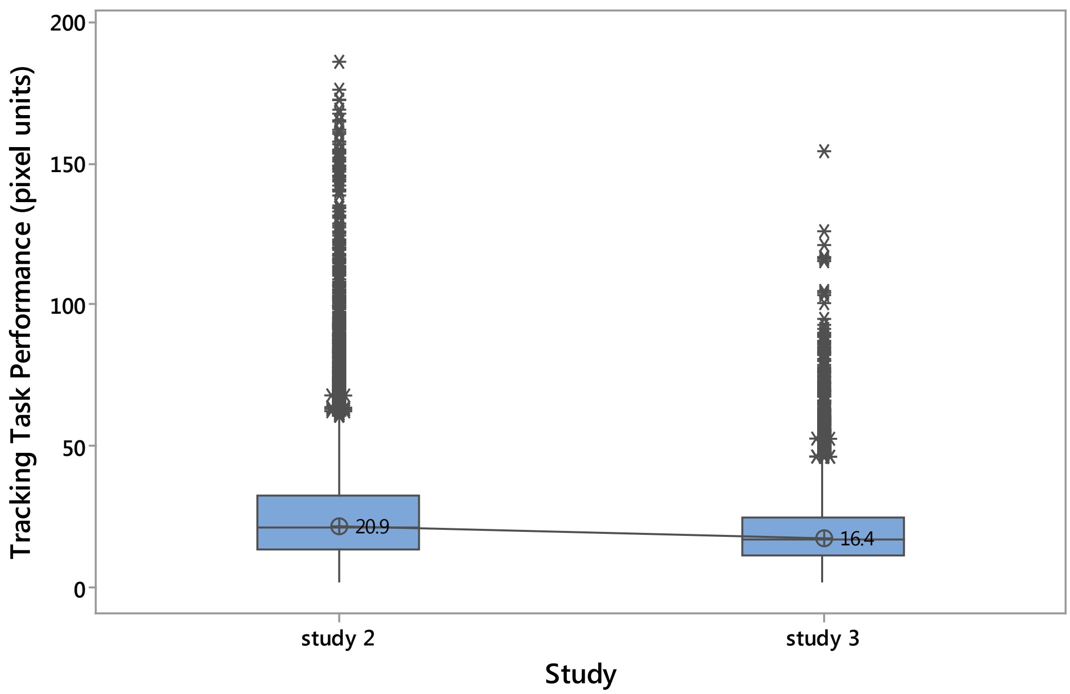


Figure 6.3 Tracking task performance (RMSD): study 2 vs. study 3

Figure 6.3 also indicates that there is the improvement in performance between study 3 and study 2.

6.4.3 Hypothesis 3b

Different combinations of threshold values were applied to the tracking task performance data collected in study 3 to see how the PB detection method performs on the data where PB did not occur. (See Appendix D.) Next, the threshold values, i.e., selected based on criterion 1 from the first study, were applied to the tracking task performance data collected in the third study. The following table (Table 6.2) reports the results.

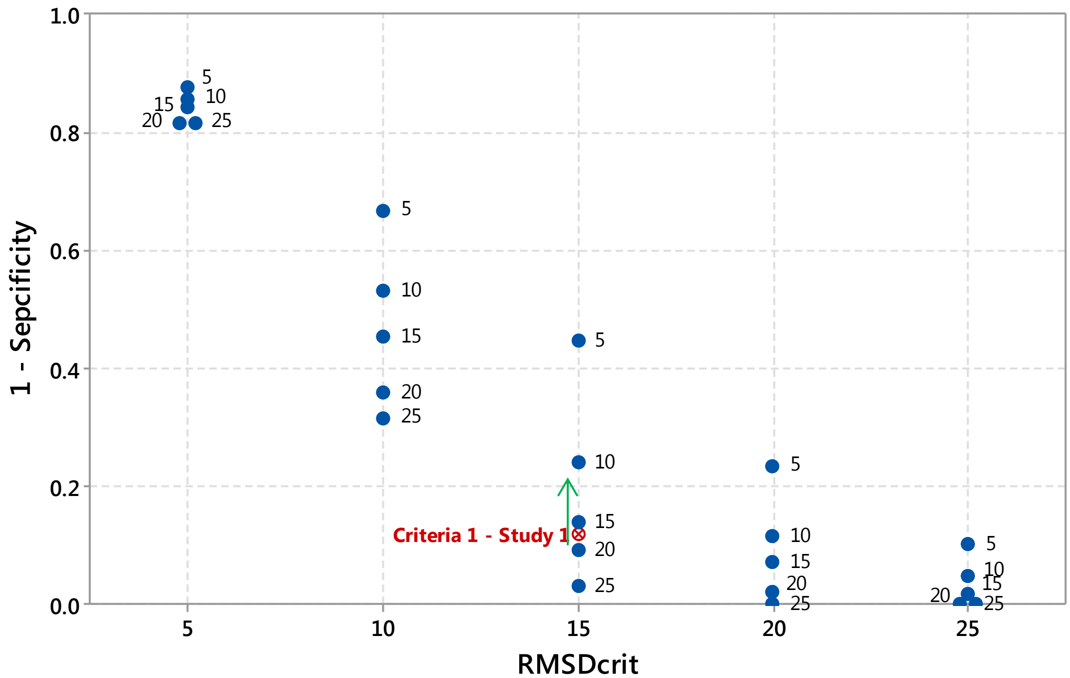
Table 6.2 The performance of re-using the threshold values selected (criterion 1) from study 1 on the data collected in study 3

Participan t	$RMSD_{crit}$	ε (seconds)	Duration of False detection (seconds)	1-Specificity (study 3)	Sensitivity (study 3)
1	15.0	10.0	133.0	0.1	0.1
2	10.0	10.0	378.0	0.1	0.4
3			No report of PB		
4	10.0	25.0	507.0	0.3	0.5
5	15.0	10.0	308.0	0.5	0.2
6	25.0	25.0	0.0	0.1	0.0
7	15.0	10.0	139.0	0.5	0.1
8			No report of PB		
9			No report of PB		
10	25.0	10.0	103.0	0.3	0.1
11	10.0	25.0	364.0	0.1	0.3
12	20.0	15.0	40.0	0.1	0.0
13	10.0	15.0	480.0	0.4	0.4
14	20.0	5.0	119.0	0.1	0.1
15	10.0	15.0	491.0	0.3	0.5

PB was identified using the same parameters of the PB detection method identified from the previous study based on criterion 1.

The following graph (Figure 6.4) shows that there is decrease in specificity when the identified threshold parameters from study 1 were re-used. The blue dots indicate 1 - specificity that were identified for each different combination of threshold parameters in study 3. The value on the top of each blue dot indicates the value of ε. The rest of the figures are included in Appendix A.

Figure 6.4 Evaluation of threshold values selected (criterion 1 and 2) from study 1 on participant 5's tracking task performance in study 3



6.5 Discussion

The same participants from study 2 participated in this study. The results indicate that none of the participants experienced self-reported PB during study 3. Such results promise a potential benefit that can be achieved by dynamically allocating secondary tasks to the automation when PB is detected. There are functions that have to be performed by a human agent, which could be assigned as a primary task. For instance, in air traffic management, the routes that the arriving aircraft have to fly are often blocked by severe weather (e.g., thunderstorms). There are limitations of automation in effectively rerouting aircraft to avoid the severe weather. These functions cannot be fully automated due to uncertainty and require human involvement particularly in the decision making process. In such conditions, a dynamic function allocation system could intervene to assist the human operator by reallocoting secondary tasks to automation.

It was identified that there was a significant difference between performance in study 2 and 3. This suggests that performance changes in study 1 and 2 after PB is not only due to the task difficulty changes but also due to the presence of the secondary task.

As the performance data collected while conducting study 3 contains no self-report of PB, the data was used to test the reliability of the PB detection method. The threshold values that were found based on criterion 1 and 2 from study 1 were examined. As no participants identified to experience PB, no PB should be ideally detected. When the PB detection method was applied to the performance data collected during study 3, there were durations of time that were identified as PB. The result suggests that the calibrated threshold value from study 1 cannot be re-used in the subsequent study.

CHAPTER 7. CONCLUSION

7.1 Conclusions

After conducting the studies, the following conclusions are made.

- Increasing workload can induce subjectively-identified PB, although it might not be generalizable.
- There exists criteria such that identifies the threshold parameters of the PB detection method that best captures the performance characteristics at the subjectively-identified PB point, however, there was no such criteria that is consistent among participants, which suggests the parameters of the PB detection method may have to be calibrated each time.
- There does not appear to be any evidence that pulse rate can be used to detect PB.
- PB is induced by the primary task workload and is affected by the presence of the secondary tasks, which suggests PB detection method may perform more effectively in a single task environment.

7.2 Contributions

The contributions of this dissertation work were made centrally filling the following gaps:

- PB has been only anecdotally described as a state where the operator "loses control of the context" and "cannot maintain the task performance."
- The past work on PB descriptions do not have specific definitions.
- PB has not been empirically demonstrated.
- There is no validated objective way of detecting PB or the transition into such state.

In this dissertation, a definition of PB is given. PB was successfully induced in a controlled setting. The criteria from the PB definition detected PB and it was shown that increasing workload can induce subjectively-identified PB, although this might not be generalizable. This suggests that the parameters of the PB detection method may have to be calibrated per individual. The parameters of the PB detection method were calibrated to objectively capture the performance characteristics when PB was subjectively indicated. Then, the evaluation was conducted to determine whether such calibrated parameters could be re-used over time. It was found that the performance characteristics at subjectively identified PB point were varied over time. Currently, the only available way of identifying PB is through subjective identification. However, possible ambiguity issues with such subjectively declared PB points were found.

7.3 Future Work

Based on the lessons learned from conducting this dissertation work, future studies can follow to extend our understanding of PB. The findings from this study suggest the need for further investigation on evaluating and improving the PB detection method.

- The parameters of the PB detection method were calibrated to match the subjectively declared PB point. There are ambiguity issues with the subjectively declared PB point. Hence, other indicators of PB in other measures should be investigated. The redundancy that could potentially be provided by multiple indicators could help by improving the reliability of PB detection.
- Other physiological measures, such as EEG, could be tested. Tools such as eye trackers could help in exploring shifts in performance strategy after PB.
- In order to prevent operators from experiencing PB, effort should be made to look for reliable precursors to PB. Such indications can be used to preemptively prevent PB from occurring in advance.
- In this dissertation work, the studies are conducted in an environment where task difficulty is only increasing. Detection of PB could be tested in an environment where task difficulty is dynamically changes throughout the experiment. This could help the researchers see how such a detection method could be used operationally in a real environment.

- PB was induced in a multi-task environment, and the presence of the secondary task clearly affected PB. Attempts to induce PB should also be conducted in a single task environment. If PB can be induced in such environment, the characteristics of PB in such environment should be reported and the PB detection method should be examined in the data.
- The PB detection method should be examined on the data collected from participants who are not only trained to be familiar with the task but also trained sufficiently to a criterion so that they become consistent in their performance between studies and minimize possible learning effect.
- There is a need for collecting human performance data that includes not only the output of the task but the input the participants provide to the task. For instance, information on when participants are moving the joystick in the tracking task, in addition to data on target movement, can help decipher PB characteristics in the data. Such information can help identifying whether the target movement was due to change that participants make or due to how the task is designed.
- In this dissertation work, two factors, the sensitivity of the joystick and the magnitude of the displacement of the target, were manipulated to increase the difficulty of the tracking task. Follow-up studies could be designed in a way that there is only one factor manipulated either sensitivity of the joystick or target movement, in order to see the effect of each manipulation on PB more clearly.

REFERENCES

REFERENCES

Acharya, U. R., Joseph, K. P., Kannathal, N., Lim, C. M., & Suri, J. S. (2006). Heart rate variability: a review. *Medical and biological engineering and computing*, 44, 1031-1051.

Andre, A.D. (2001). The value of workload in the design and analysis of consumer products. In P.A.Hancock, & P.A. Desmond, *Stress, Workload, and Fatigue,* 373–382. Mahwah, NJ: L.Erlbaum.

Backs, R. W., Lenneman, J. K., Wetzel, J. M., & Green, P. (2003). Cardiac measures of driver workload during simulated driving with and without visual occlusion. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 45, 525-538.

Bailey, N. R., Scerbo, M. W., Freeman, F. G., Mikulka, P. J., & Scott, L. A. (2006). Comparison of a brain-based adaptive system and a manual adaptable system for invoking automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 48, 693-709.

Barnes, M., & Grossman, J. (1985). The intelligent assistant for electronic warfare systems (Rep. No. NWC TP 5885) China Lake, CA: U.S. Naval Weapons Center.

Billings, C. (1991). Human-centered aircraft automation: A concept and guidelines. (NASA Tech. Memorandum103885).Moffett Field, CA: NASA Ames Research Center.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30, 1145-1159.

Broadbent, D. E. (1958). *Perception and communication*. New York: Permagon Press.

Brown, S. W. (1997). Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. *Perception & Psychophysics*, 59, 1118-1140.

Byrne, E. A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biological Psychology,* 42, 249-268.

Caldwell, B. S. (1997). Components of information flow to support coordinated task performance. *International Journal of Cognitive Ergonomics*, 1, 25-41.

Caldwell, B., & Onken, J. (2011). Modeling and analyzing distributed autonomy for spaceflight teams. In *Proceedings of International Conference on Environmental Systems.*

Carpenter, P. A., Just, M. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1999). Graded functional activation in the visuo-spatial system with the amount of task demand. *Journal of Cognitive Neuroscience*, 11, 9-24.

Chen, J. Y., Drexler, J. M., Sciarini, L. W., Cosenzo, K. A., Barnes, M. J., & Nicholson, D. (2008). *Operator Workload and Heart-Rate Variability During a Simulated Reconnaisance Mission with an Unmanned Ground Vehicle*. Army research lab, Orlando, FL, Human Research and Engineering Directorate.

Comstock, J.R. & Arnegard, R.J. (1992). The multi-attribute task battery for human operator workload and strategic behavior research. NASA Technical Memorandum No. 104174.

Craik, K. J. (1948). Theory of the human operator in control systems. *British Journal of Psychology. General Section*, 38, 142-148.

Dell'Erba, G., Pancheri, P., & Intreccialagli, B. (1988). Hormonal assessment and workload correlates in air traffic controllers at the end of night shift: The stress perspective. *New Trends in Experimental & Clinical Psychiatry*, 59, 251-256.

Erzberger, H. (2004). Transforming the NAS: The next generation air traffic control system. In *24th International Congress of the Aeronautical Sciences, Yokohama, Japan*.

Farley, T., Foster, J. D., Hoang, T., & Lee, K. K. (2001). A time-based approach to metering arrival traffic to Philadelphia. In *Proceedings of First FIAA Aircraft Technology, Integration, and Operations Forum,* Los Angeles, California.

Feigh, K. M., Pritchett, A. R., Jacko, J. A., & Denq, T. (2005). Decision making during an airline rescheduling task: a contextual control model description. In *Proceedings of the International symposium on Aviation Psychology,* Oklahoma City, OK.

Gevins, A., Smith, M. E., McEvoy, L., & Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. *Cerebral Cortex*, 7, 374-385.

Gjerde, P. F. (1983). Attentional capacity dysfunction and arousal in schizophrenia. *Psychological Bulletin*, 93, 57.

Gopher, D., & Donchin, E. (1986). Workload-An examination of the concept. In K. Boff, l. Kauf-man, & J. Thomas (Eds.), *Handbook of perception and performance: Cognitive processes and performance*. New York: Wiley

Hancock, P. A., & Ganey, H. C. (2003). From the inverted-U to the extended-U: The evolution of a law of psychology. *Journal of Human Performance in Extreme Environments*, 7, 3.

Haarmann, A., Boucsein, W., & Schaefer, F. (2009). Combining electrodermal responses and cardiovascular measures for probing adaptive automation during simulated flight. *Applied Ergonomics*, 40, 1026-1040.

Hilburn, B., Jorna, P. G., Byrne, E. A., & Parasuraman, R. (1997). The effect of adaptive air traffic control (ATC) decision aiding on controller mental workload. In M. Mouloua & J. Koonce (Eds.), *Human-Automation Interaction*. Mahwah, NJ: Erlbaum.

Hilburn, B., & Jorna, P. G. A. M. (2001). Workload and air traffic control. *Stress, Workload and Fatigue: Theory, Research and Practice.* Hillsdale, New Jersey, USA: Erlbaum.

Hockey, G. R. J. (1986). A state control theory of adaptation and individual differences in stress management. In *Energetics and Human Information Processing*, 285-298.

Hockey, G. R. J. (2003). *Operator functional state: the assessment and prediction of human performance degradation in complex tasks*, IOS Press.

Hoffman, J. E., Nelson, B., & Houck, M. R. (1983). The role of attentional resources in automatic detection. *Cognitive Psychology*, 15, 379-410.

Hollnagel, E. (1993). Models of cognition: procedural prototypes and contextual control. *Le travail humain*, 27-51.

Inagaki, T. (2003). Adaptive automation: Sharing and trading of control. In E. Hollnagel (Ed.), *Handbook of Cognitive Task Design*, 147-169. Mahwah, N.J.: Lawrence Erlbaum Pub.

Jorna, P. G. A. M. (1993). Heart rate and workload variations in actual and simulated flight. *Ergonomics*, 36, 1043-1054.

Kaber, D. B., & Endsley, M. R. (1997). The combined effect of level of automation and adaptive automation on human performance with complex, dynamic control systems. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 41,205-209. SAGE Publications.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliff, NJ: Prentice-Hall.

Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An efficient k-means clustering algorithm: Analysis and implementation. *Pattern Analysis and Machine Intelligence, IEEE Transactions*, 24, 881-892.

Kerstholt, J. (1996). The effect of information costs on strategy selection in dynamic tasks. *Acta Psychologica*, 94, 273-290.

Kuchar, J. K. (1996). Methodology for alerting-system performance evaluation. *Journal of Guidance, Control, and Dynamics*, 19, 438-444.

Lagu, A. V. (2009). *Adaptive function allocation for human-machine systems: A comparative study of triggers, strategies and stabilization.* PhD Dissertation, Purdue University.

Lagu, A. V., & Landry, S. J. (2011). Roadmap for the next generation of dynamic function allocation theories and strategies. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 21, 14-28.

Lagu, A. V., Landry, S. J., & Yoo, H. S. (2013). Adaptive function allocation stabilization and a comparison of trigger types and adaptation strategies. *International Journal of Industrial Ergonomics*.

Landry, S. J., & Yoo, H. S. (2009). Modeling Operator Responses to Alerts. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 53, 41-45.

Landry, S. J. (2012). The next generation air transportation system: An approach to function allocation. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 22, 351-361.

Leiden, K., Laughery, K. R., Keller, J., French, J., Warwick, W., & Wood, S. D. (2001). A review of human performance models for the prediction of human error. *NASA system-wide accident prevention program, Ames Research Center.*

Lewis, M. J. (2005). Heart rate variability analysis: a tool to assess cardiac autonomic function. *Computers Informatics Nursing*, 23, 335-341.

Li, H., Sarter, N., Wickens, C., & Sebok, A. (2013). Supporting Human-Automation Collaboration Through Dynamic Function Allocation The Case of Space Teleoperation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*.

Light, K. C., & Obrist, P. A. (1983). Task difficulty, heart rate reactivity, and cardiovascular responses to an appetitive reaction time task. *Psychophysiology*, 20, 301-312.

Liu, Y. (1996). Queueing network modeling of elementary mental processes. *Psychological Review*, 103, 116.

Liu, Y. (1997). Queueing network modeling of human performance of concurrent spatial and verbal tasks. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions,* 27, 195-207.

Maule, A. J. (1997). Strategies for adapting to time pressure. *Decision making under stress- Emerging themes and applications(A 99-12526 01-53), Aldershot, United Kingdom, Ashgate,* 271-279.

McClain, E., Clarke, J. P., Huang, A., & Schleicher, D. (2009). Traffic volume intersection metric for metroplex clustering analysis. In *AIAA Guidance, Navigation, and Control Conference*, 10-13.

Mikulka, P. J., Scerbo, M. W., & Freeman, F. G. (2002). Effects of a biocybernetic system on vigilance performance. *Human Factors,* 44, 654-664.

Miller, C.,& Parasuraman, R. (2007). Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. *Human Factors*, 49, 57–75

Molloy, R., & Parasuraman, R. (1996). Monitoring an automated system for a single failure: Vigilance and task complexity effects. *Human Factors*, 38, 311-322.

Navon, D., & Miller, J. (1987). Role of outcome conflict in dual-task interference. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 435.

Neiss, R. (1988). Reconceptualizing arousal: psychobiological states in motor performance. *Psychological Bulletin*, 103, 345.

Parasuraman, R., Bahri, T., Deaton, J. E., Morrison, J. G., & Barnes, M. (1992). *Theory and design of adaptive automation in aviation systems*. (Progress Rep No. NAWCADWAR-92033-60). Warminster, PA: Naval Air Warfare Center.

Parasuraman, R., Cosenzo, K. A., & De Visser, E. (2009). Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload. *Military Psychology,* 21, 270-297.

Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors*, 38, 665-679.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors,* 39, 230-253.

Parasuraman, R., Sheridan, T., & Wickens, C. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans,* 30*,* 286–297.

Parshuram, C. S., Dhanani, S., Kirsh, J. A., & Cox, P. N. (2004). Fellowship training, workload, fatigue and physical stress: a prospective observational study. *Canadian Medical Association Journal*, 170, 965-970.

Prevot, T., Homola, J., & Mercer, J. (2008). Initial study of controller/automation integration for NextGen separation assurance. In *AIAA Guidance, Navigation, and Control (GNC) Conference and Exhibit*.

Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., & Pope, A. T. (2000). A closed-loop system for examining psychophysiological measures for adaptive task allocation. *The International Journal of Aviation Psychology*, 10, 393-410.

Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., & Pope, A. T. (2003). Effects of a psychophysiological system for adaptive automation on performance, workload, and the event-related potential P300 component. *Human Factors,* 45, 601-614.

Rasmussen, J. (1997). Risk management in a dynamic society: a modelling problem. *Safety Science*, 27, 183-213.

Recarte, M. A., & Nunes, L. M. (2000). Effects of verbal and spatial-imagery tasks on eye fixations while driving. *Journal of Experimental Psychology: Applied,* 6, 31.

Roscoe, A. H. (1993). Heart rate as a psychophysiological measure for in-flight workload assessment. *Ergonomics*, 36, 1055-1062.

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey & E. Salas (Eds.), *Teams: Their training and performance*. Norwood, NJ: Ablex.

Saraf, A., Clarke, J. P., & McClain, E. (2010). Discussion and Comparison of Metroplex-wide Arrival Scheduling Algorithms. In *AIAA Aviation, Technology, Integration, and Operations Conference*.

Sauer, J., Kao, C. S., Wastell, D., & Nickel, P. (2011). Explicit control of adaptive automation under different levels of environmental stress. *Ergonomics,* 54, 755-766.

Scerbo, M. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and Applications*, 37–63, Mahwah, NJ: Erlbaum.

Sheridan, T. B., & Verplanck, W. L. (1978). *Human and Computer Control of Undersea Teleoperators*. Tech. Rep., MIT Man-Machine Systems Laboratory, Cambridge, MA.

Sheridan, T. B. (2000). Function allocation: algorithm, alchemy or apostasy?. *International Journal of Human-Computer Studies*, 52, 203-216.

Sheridan, T. B. (2006). Supervisory control. *Handbook of Human Factors and Ergonomics, Third Edition*, 1025-1052.

Stanton, N. A., Ashleigh, M. J., Roberts, A. D., & Xu, F. (2001). Testing hollnagel's contextual control model: Assessing team behavior in a human supervisory control task. *International Journal of Cognitive Ergonomics*, 5, 111-123.

Swets, J. A. (2012). *Evaluation of diagnostic systems*. Elsevier.

Swets, J. A. (2014). *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers*. Psychology Press.

Ting, C. H., Mahfouf, M., Nassef, A., Linkens, D. A., Panoutsos, G., Nickel, P., & Hockey, G. (2010). Real-time adaptive automation system based on identification of operator functional state in simulated process control operations. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions*, 40, 251-262.

Todd, P. A., & Benbasat, I. (1994). The influence of decision aids on choice strategies. under conditions of high cognitive load. *Systems, Man and Cybernetics, IEEE Transactions on*, 24, 537-547.

Veltman, J. A., & Gaillard, A. W. K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics*, 41, 656-669.

Whinghter, L. J., Cunningham, C. J., Wang, M., & Burnfield, J. L. (2008). The moderating role of goal orientation in the workload-frustration relationship. *Journal of Occupational Health Psychology*, 13, 283.

Wickens, C. D. (1981). Processing Resources in Attention, Dual Task Performance, and Workload Assessment (No. EPL-81-3/ONR-81-3). *Illinois University at Urbana Engineering-Psychology Research Lab*.

Wickens, C.D. (1984). Processing resources in attention. In R. Parasuraman & D.R. Davies (Eds.), *Varieties of attention*,63–101. New York: Academic Press.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, 3, 159-177.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, 50, 449-455.

Wilson, G. F., Caldwell, J. A., & Russell, C. A. (2007). Performance and psychophysiological measures of fatigue effects on aviation related tasks of varying difficulty. *The International Journal of Aviation Psychology,* 17, 219-247.

Wilson, G. F. (1992). Applied use of cardiac and respiration measures: Practical considerations and precautions. *Biological Psychology*, 34, 163-178.

Wilson, G. F., & Russell, C. A. (2003). Operator functional state classification using multiple psychophysiological features in an air traffic control task. *Human Factors,* 45, 381-389.

Wilson, G. F., & Russell, C. A. (2007). Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding. *Human Factors,* 49, 1005-1018.

Wilson, G. M., & Sasse, M. A. (2000). Do users always know what's good for them? Utilizing physiological responses to assess media quality. In *People and Computers XIV—Usability or Else!*, 327-339. Springer London.

Wright, R. A., Contrada, R. J., & Patane, M. J. (1986). Task difficulty, cardiovascular response, and the magnitude of goal valence. *Journal of Personality and Social Psychology*, 51, 837.

Yerkes, R.M., & Dodson, J.D. (1908). The relation of strength of stimulus to rapidity of habit information. *Journal of Comparative Neurology of Psychology,* 18, 459–482.

Yoo, H. S. (2012). Framework for designing Adaptive Automation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 56, 2133-2136.

APPENDICES

Appendix A STUDY ONE (DATA)

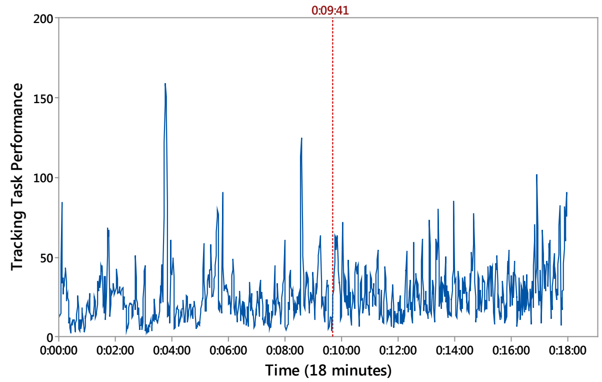


Figure A – 1 Tracking task performance (RMSD) vs. time (Participant 1)

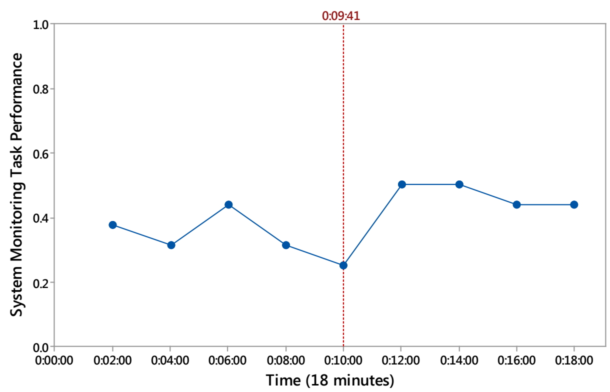


Figure A – 2 System monitoring task performance (participant 1)

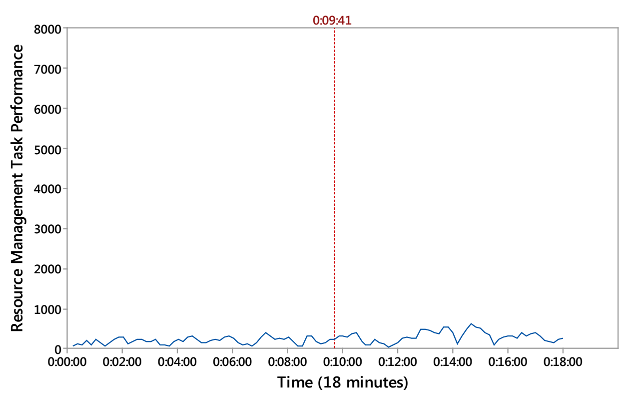


Figure A – 3 Resource management task (participant 1)

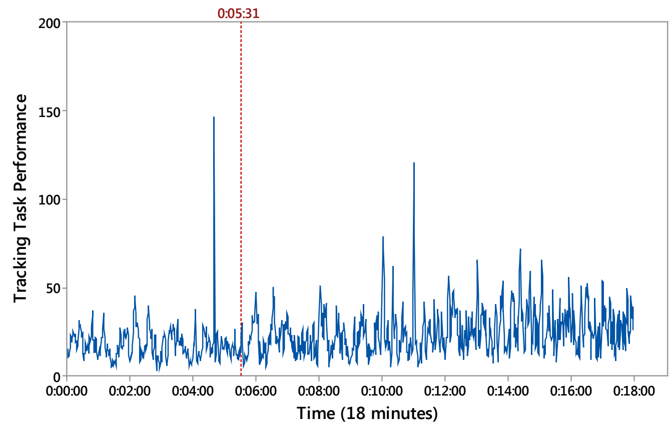


Figure A – 4 Tracking task performance (RMSD) vs. time (Participant 2)

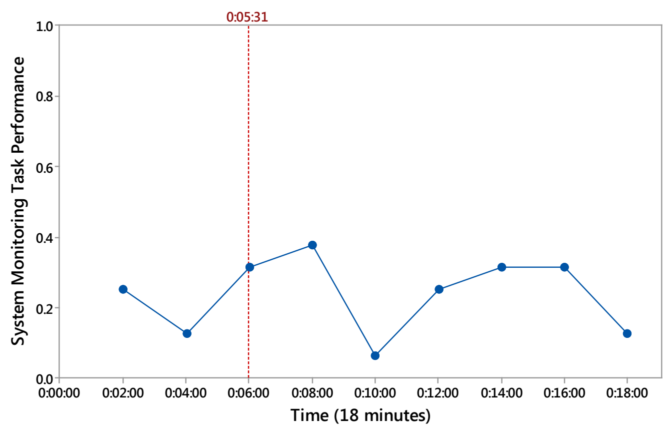


Figure A – 5 System monitoring task performance (participant 2)

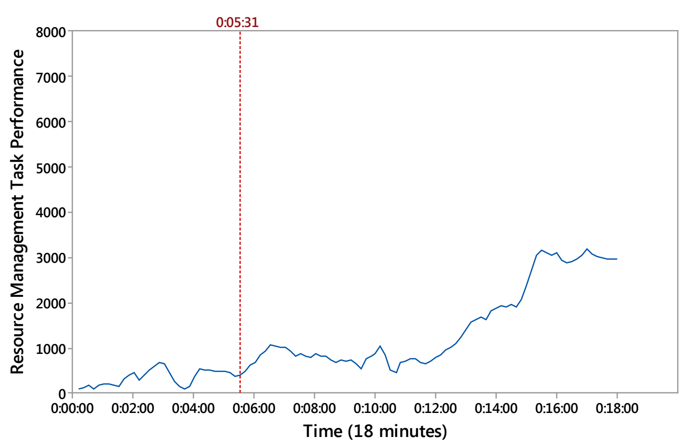


Figure A – 6 Resource management task (participant 2)

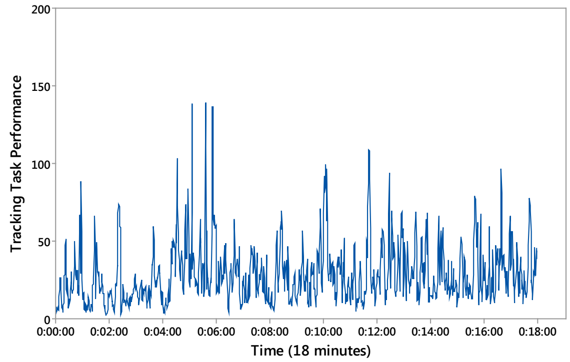


Figure A – 7 Tracking task performance (RMSD) vs. time (Participant 3)

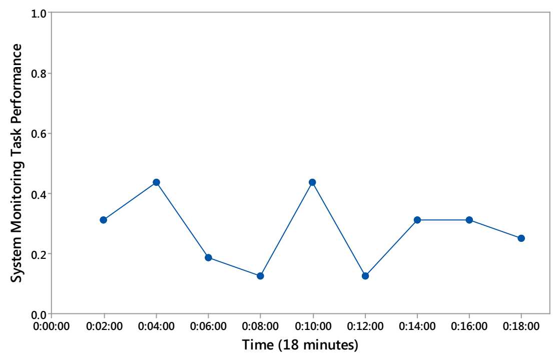


Figure A – 8 System monitoring task performance (participant 3)

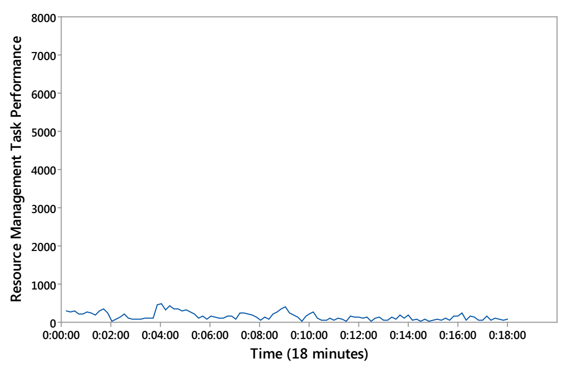


Figure A – 9 Resource management task (participant 3)

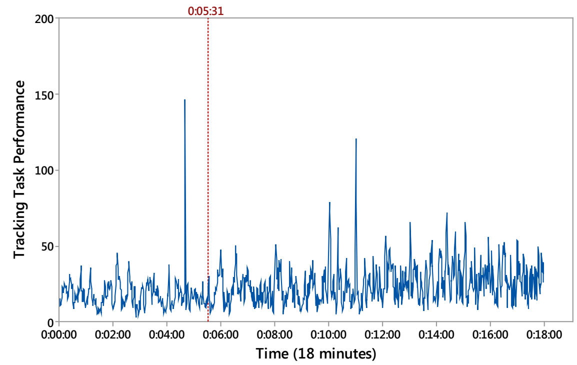


Figure A – 10 Tracking task performance (RMSD) vs. time (Participant 4)

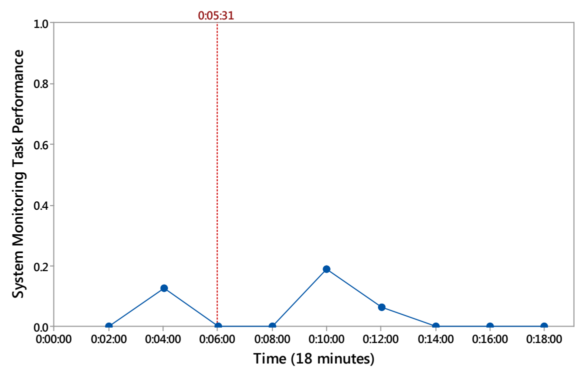


Figure A – 11 System monitoring task performance (participant 4)

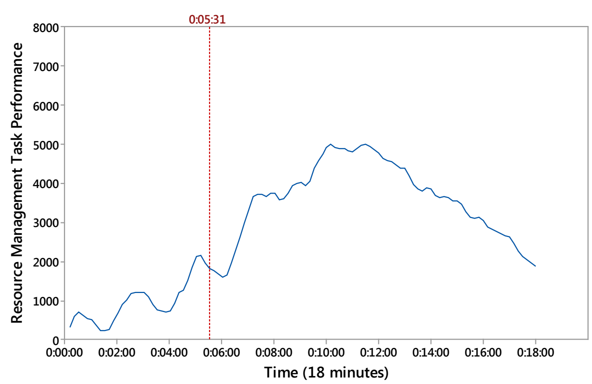


Figure A – 12 Resource management task (participant 4)

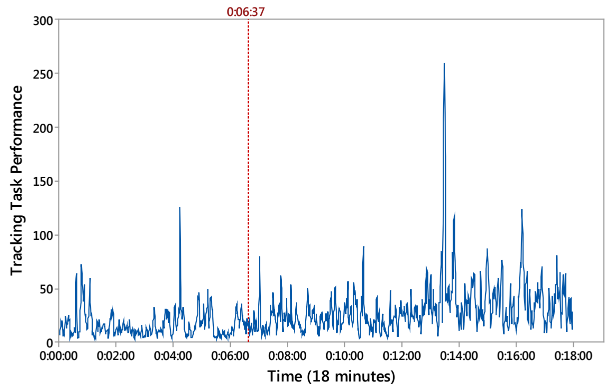


Figure A – 13 Tracking task performance (RMSD) vs. time (Participant 5)

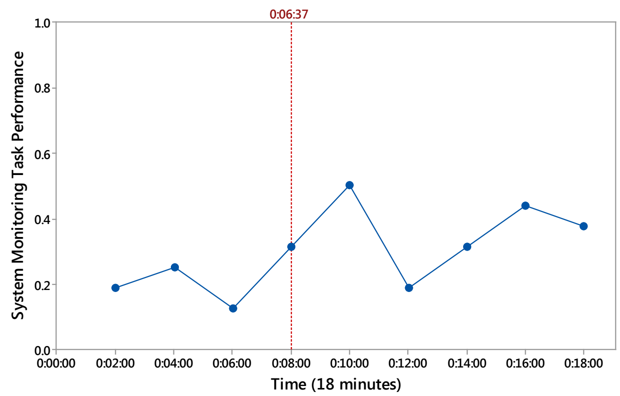


Figure A – 14 System monitoring task performance (participant 5)

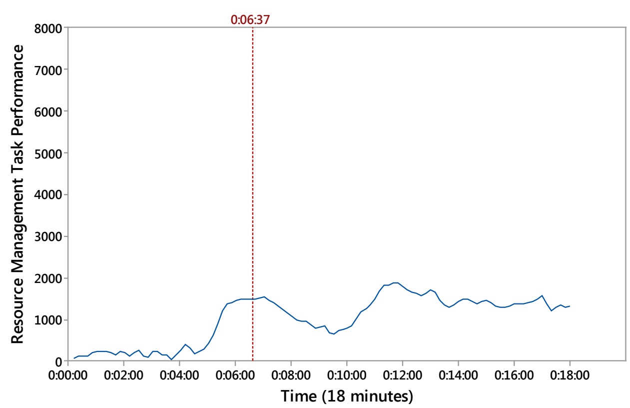


Figure A – 15 Resource management task (participant 5)

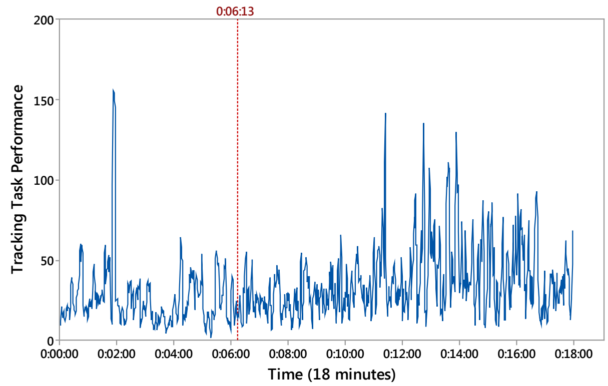


Figure A – 16 Tracking task performance (RMSD) vs. time (Participant 6)

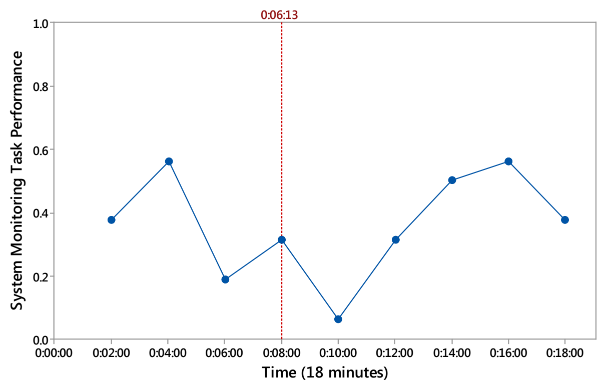


Figure A – 17 System monitoring task performance (participant 6)

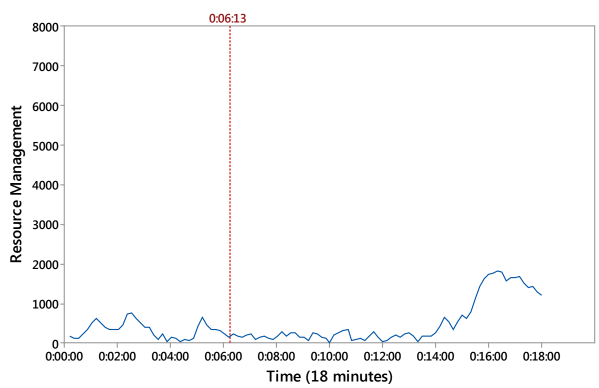


Figure A – 18 Resource management task (participant 6)

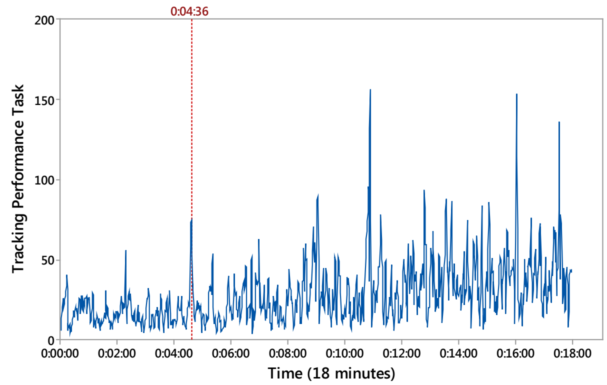


Figure A – 19 Tracking task performance (RMSD) vs. time (Participant 7)

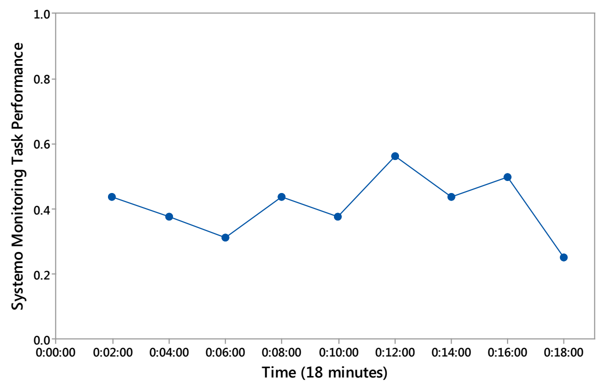


Figure A – 20 System monitoring task performance (participant 7)

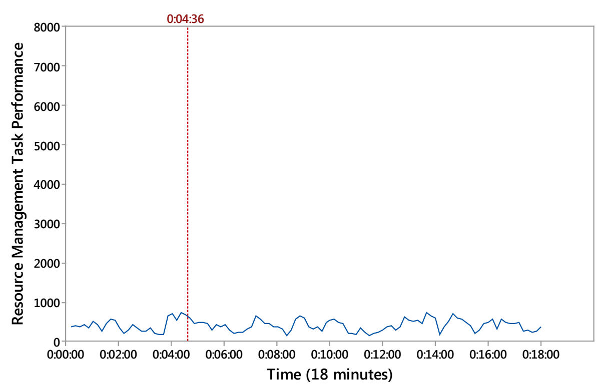


Figure A – 21 Resource management task (participant 7)

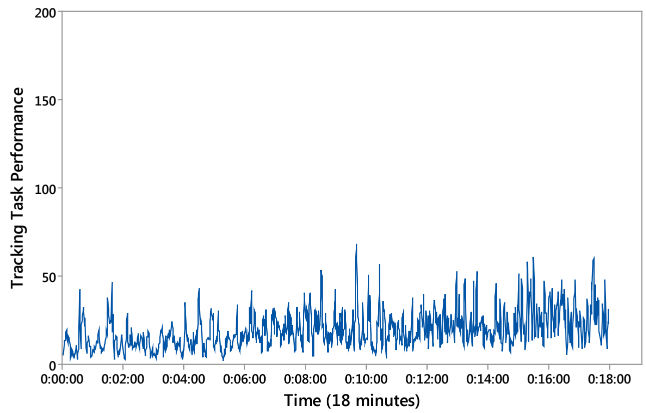


Figure A – 22 Tracking task performance (RMSD) vs. time (Participant 8)

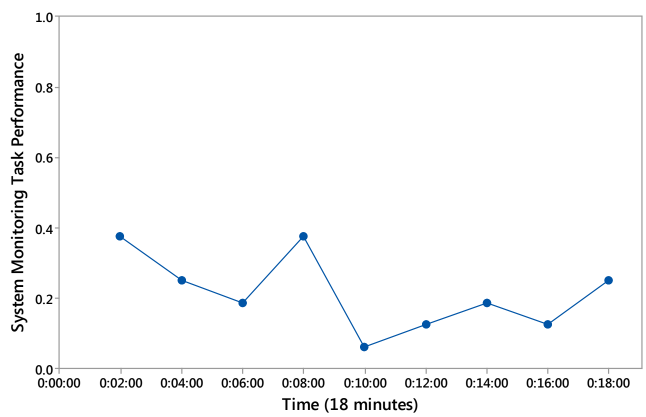


Figure A – 23 System monitoring task performance (participant 8)

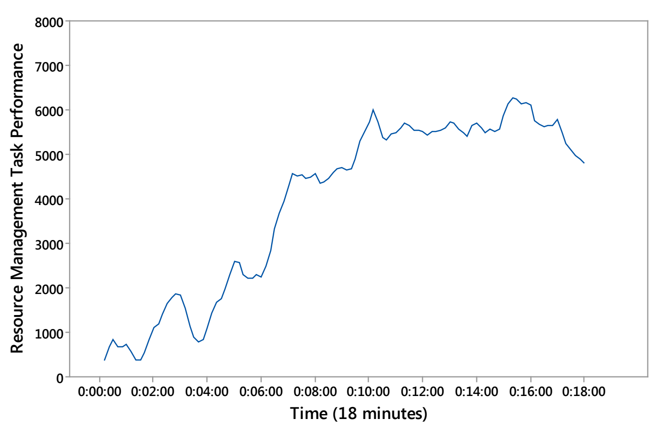


Figure A – 24 Resource management task (participant 8)

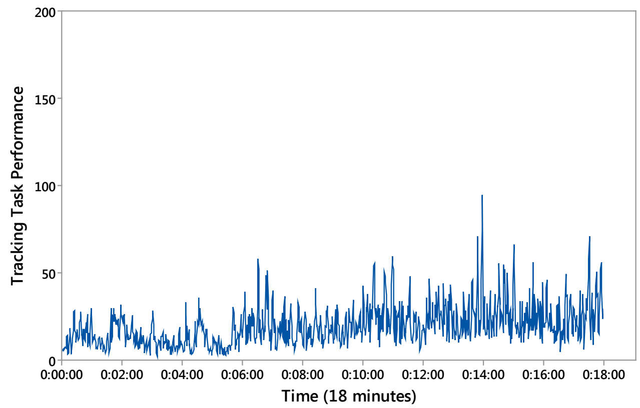


Figure A – 25 Tracking task performance (RMSD) vs. time (Participant 9)

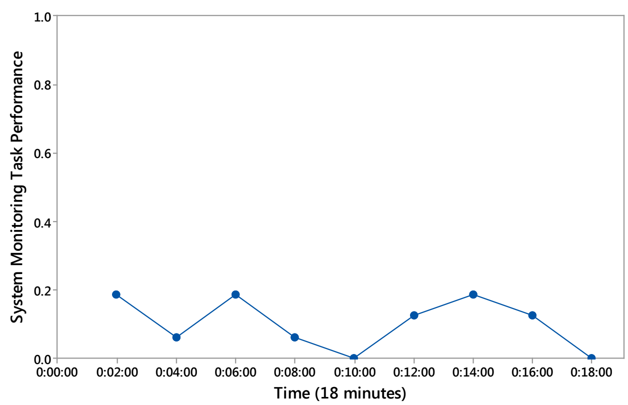


Figure A – 26 System monitoring task performance (participant 9)

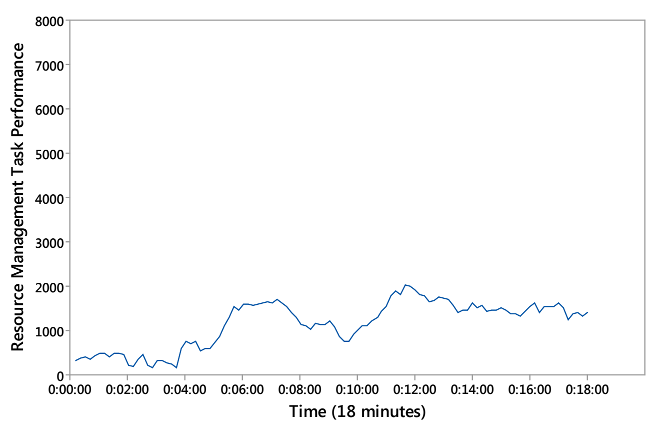


Figure A – 27 Resource management task (participant 9)

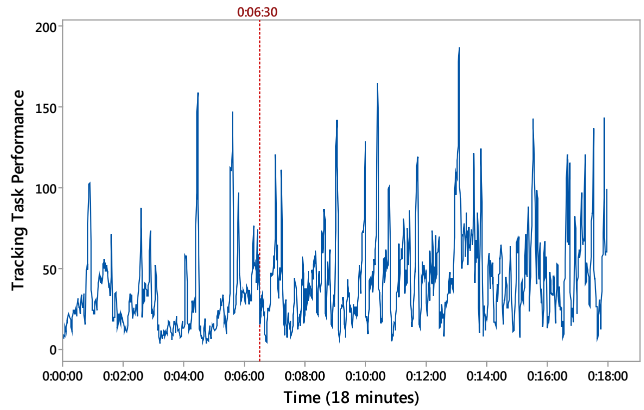


Figure A – 28 Tracking task performance (RMSD) vs. time (Participant 10)

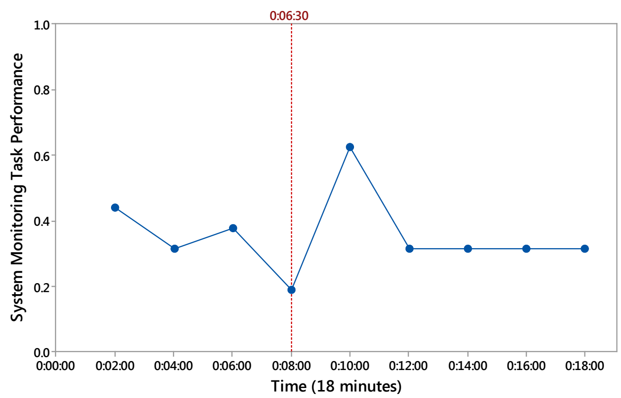


Figure A – 29 System monitoring task performance (participant 10)

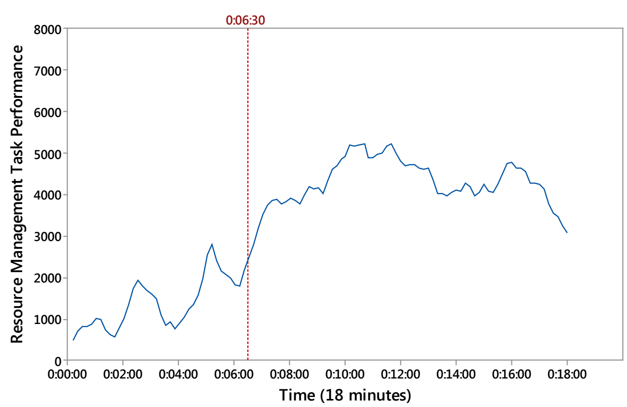


Figure A – 30 Resource management task (participant 10)

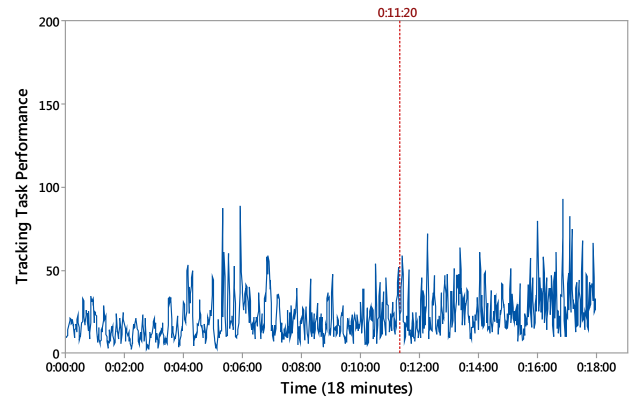


Figure A – 31 Tracking task performance (RMSD) vs. time (Participant 11)

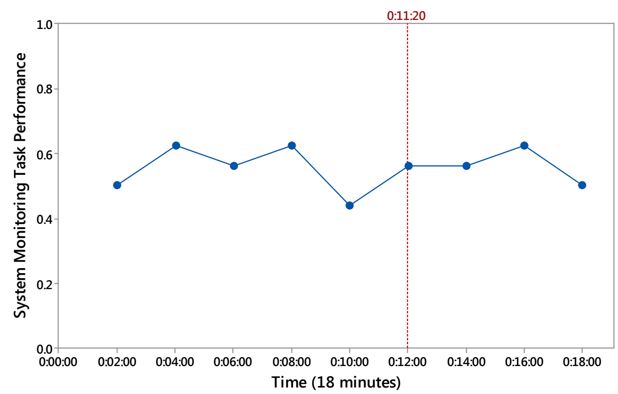


Figure A – 32 System monitoring task performance (participant 11)

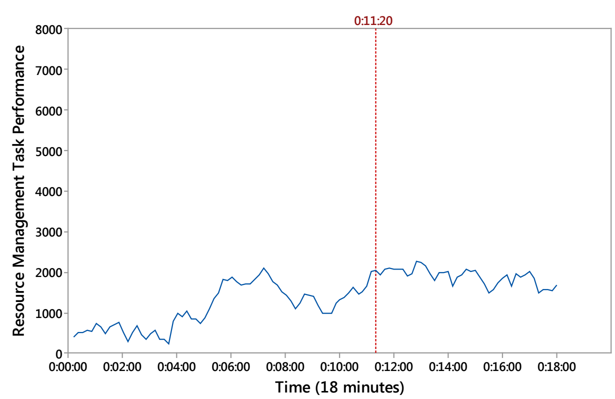


Figure A – 33 Resource management task (Participant 11)

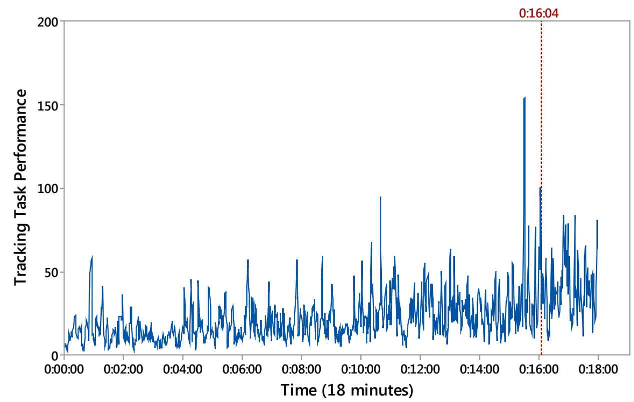


Figure A – 34 Tracking task performance (RMSD) vs. time (Participant 12)

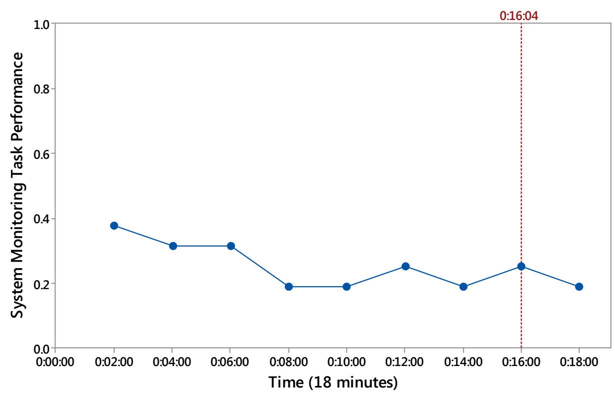


Figure A – 35 System monitoring task performance (participant 12)

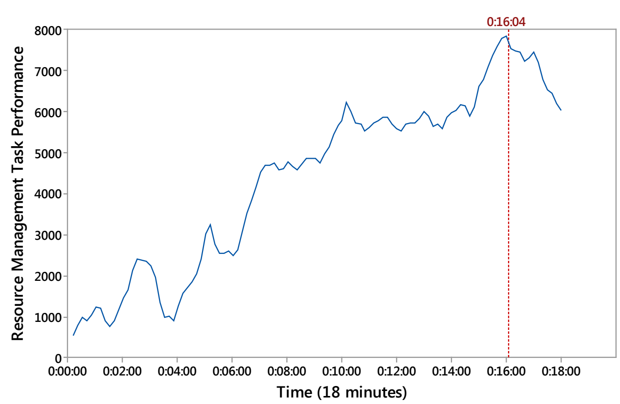


Figure A – 36 Resource management task (participant 12)

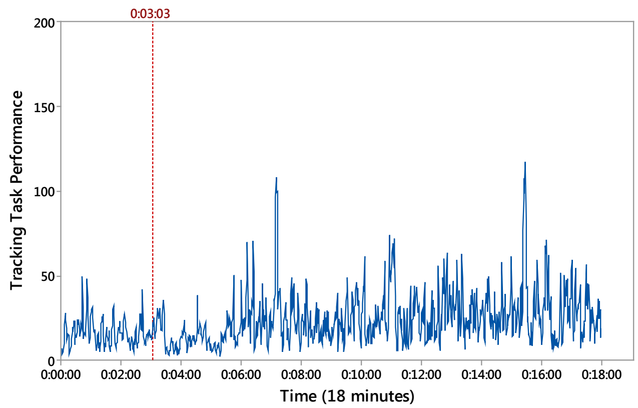


Figure A – 37 Tracking task performance (RMSD) vs. time (Participant 13)

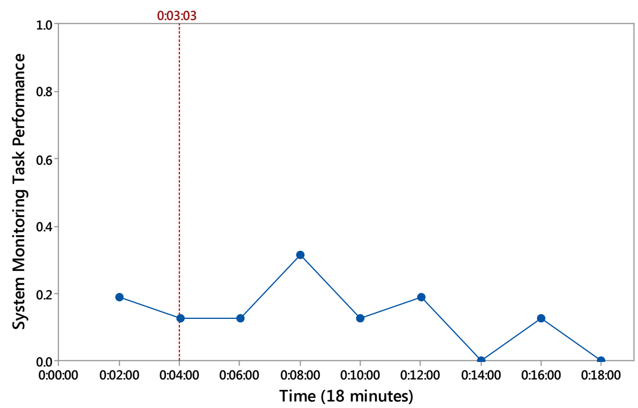


Figure A – 38 System monitoring task performance (participant 13)

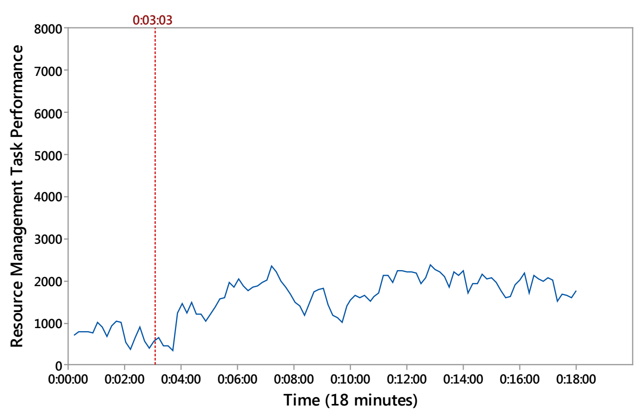


Figure A – 39 Resource management task (participant 13)

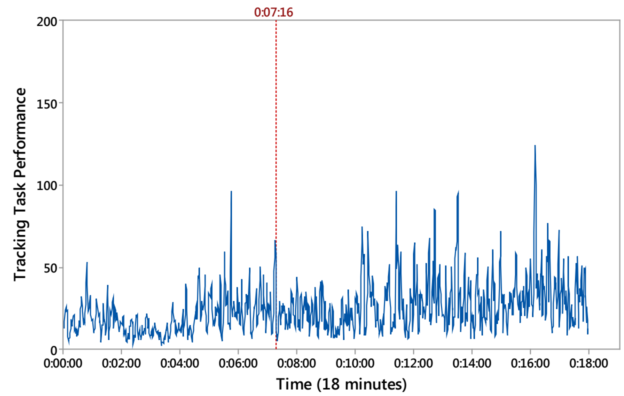


Figure A – 40 Tracking task performance (RMSD) vs. time (Participant 14)

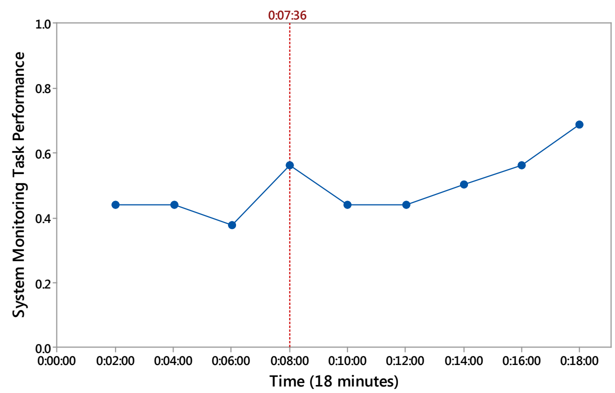


Figure A – 41 System monitoring task performance (participant 14)

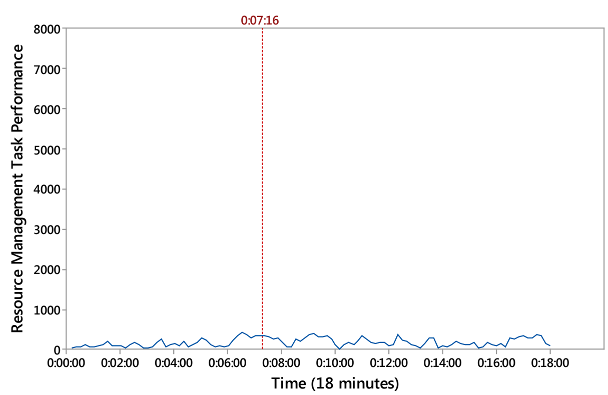


Figure A – 42 Resource management task (participant 14)

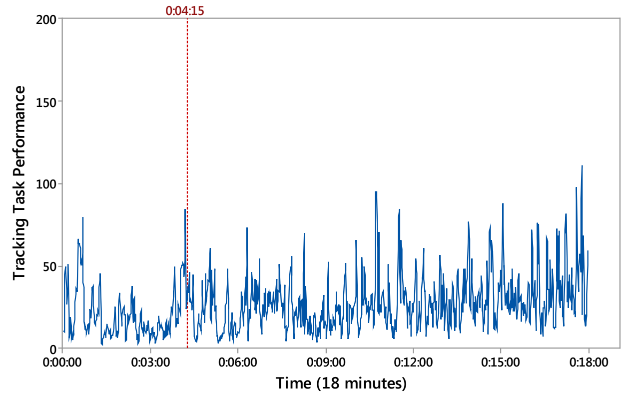


Figure A – 43 Tracking task performance (RMSD) vs. time (Participant 15)

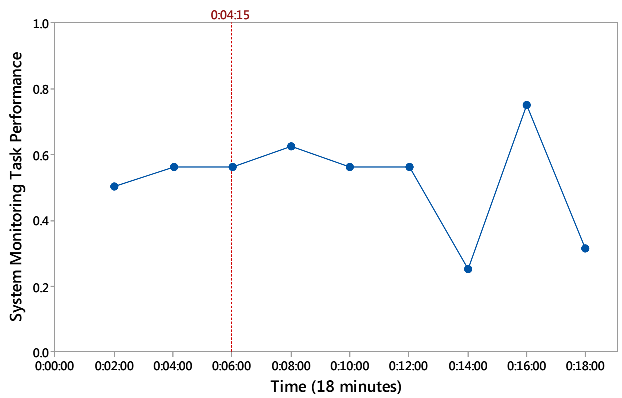


Figure A – 44 System monitoring task performance (participant 15)

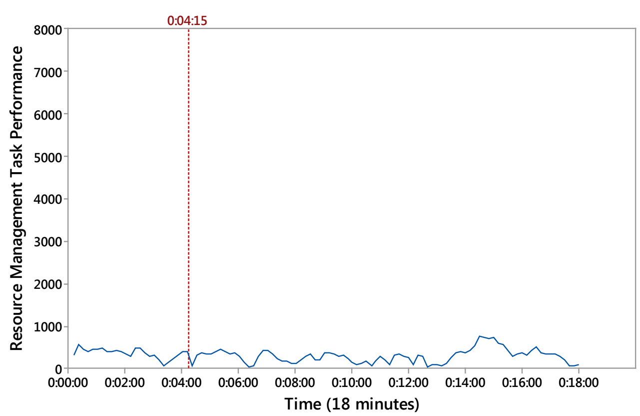


Figure A – 45 Resource management task (participant 15)

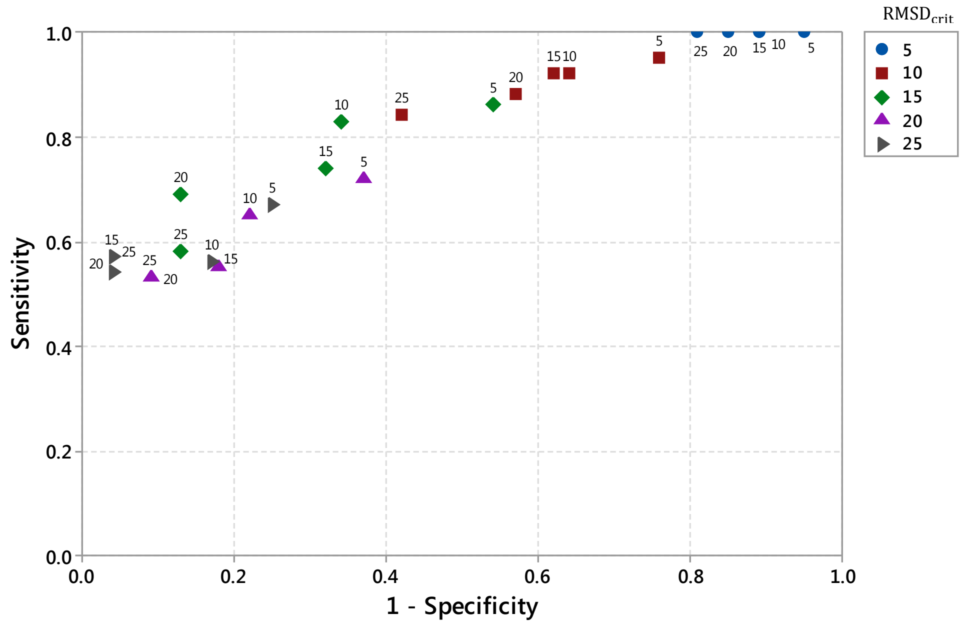


Figure A – 46 ROC curve: evaluation of parameters (RMSDcrit, ε) on participant 1's tracking task performance

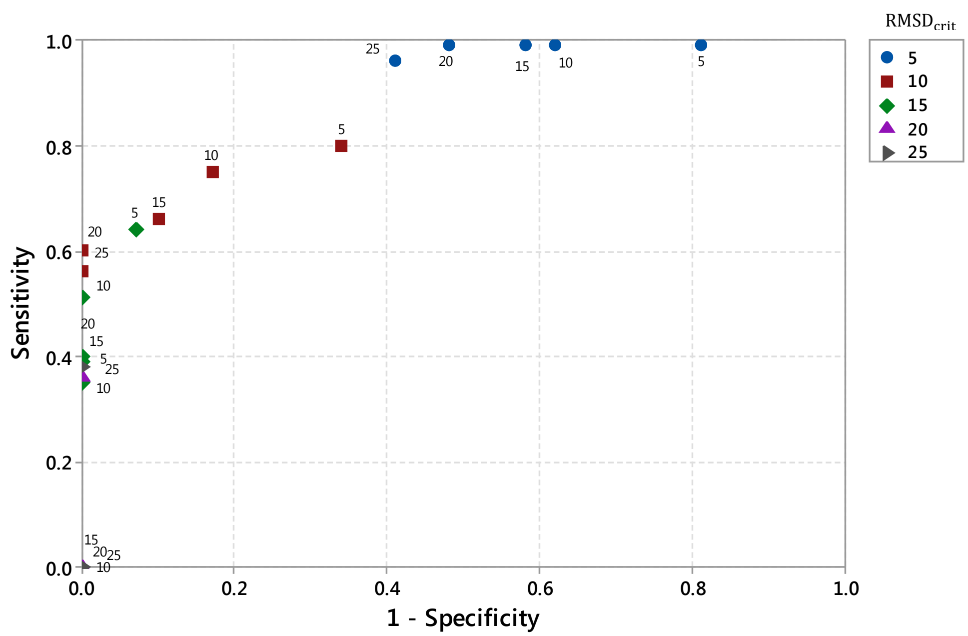


Figure A – 47 ROC curve: evaluation of parameters (RMSDcrit, ε) on participant 2's tracking task performance

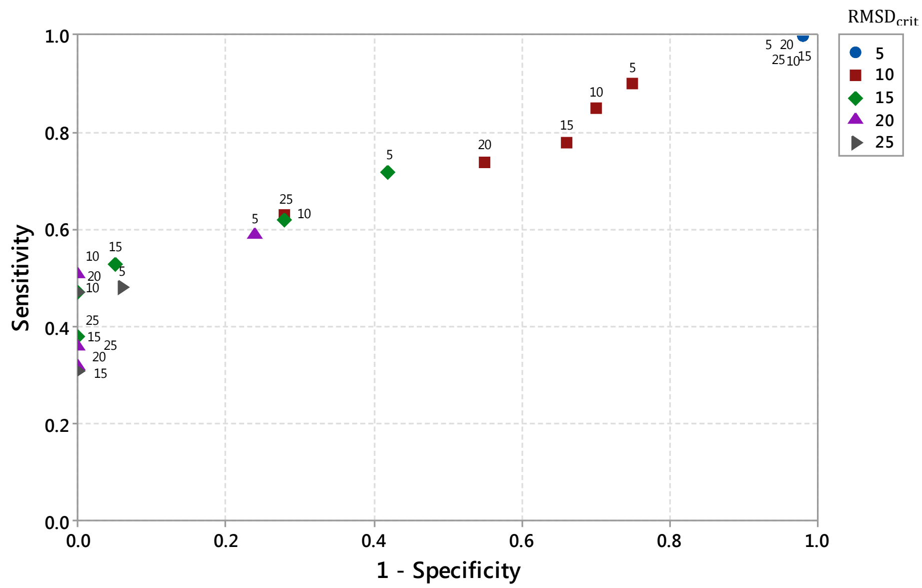


Figure A – 48 ROC curve: evaluation of parameters (RMSDcrit, ε) on participant 4's tracking task performance

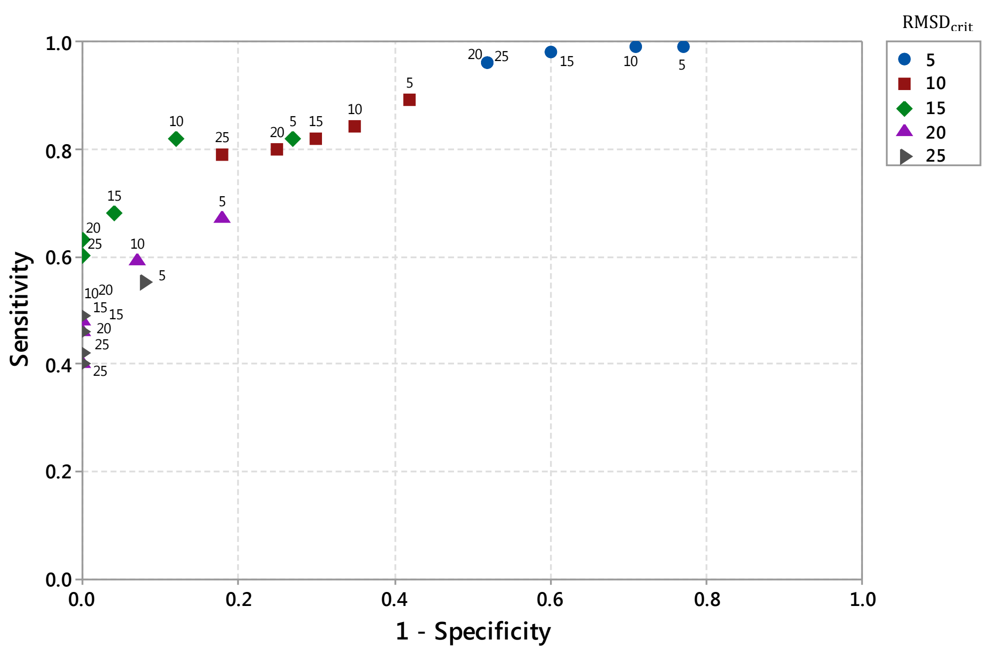


Figure A – 49 ROC curve: evaluation of parameters (RMSDcrit, ε) on participant 5's tracking task performance

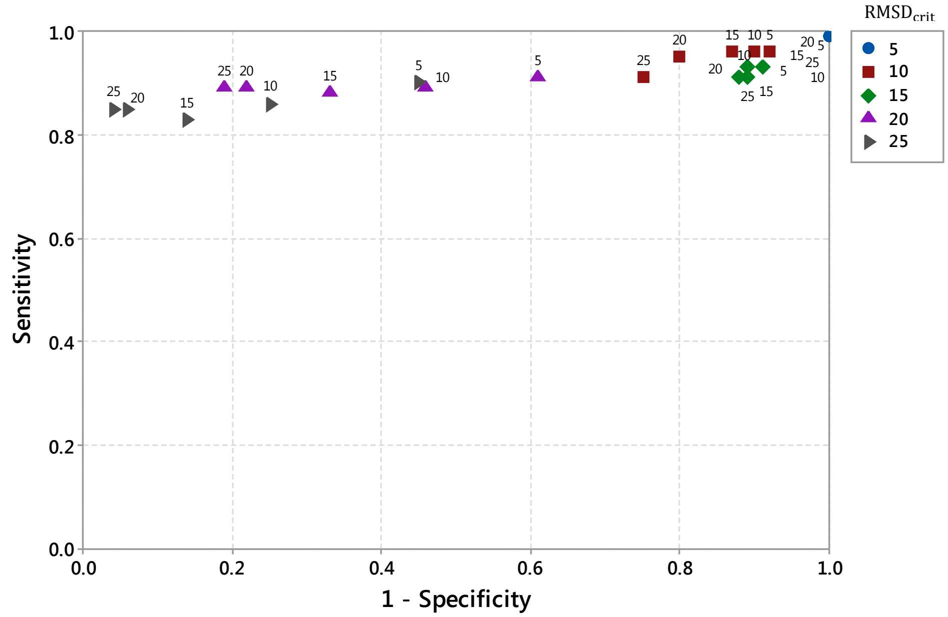


Figure A – 50 ROC curve: evaluation of parameters (RMSDcrit, ε) on participant 6's tracking task performance

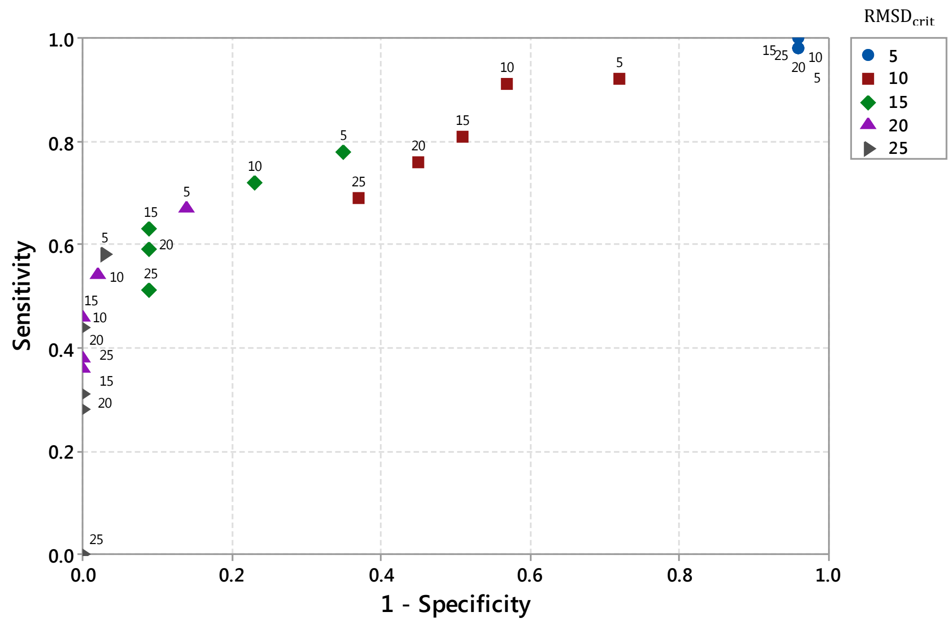


Figure A – 51 ROC curve: evaluation of parameters (RMSDcrit, ε) on participant 7's tracking task performance

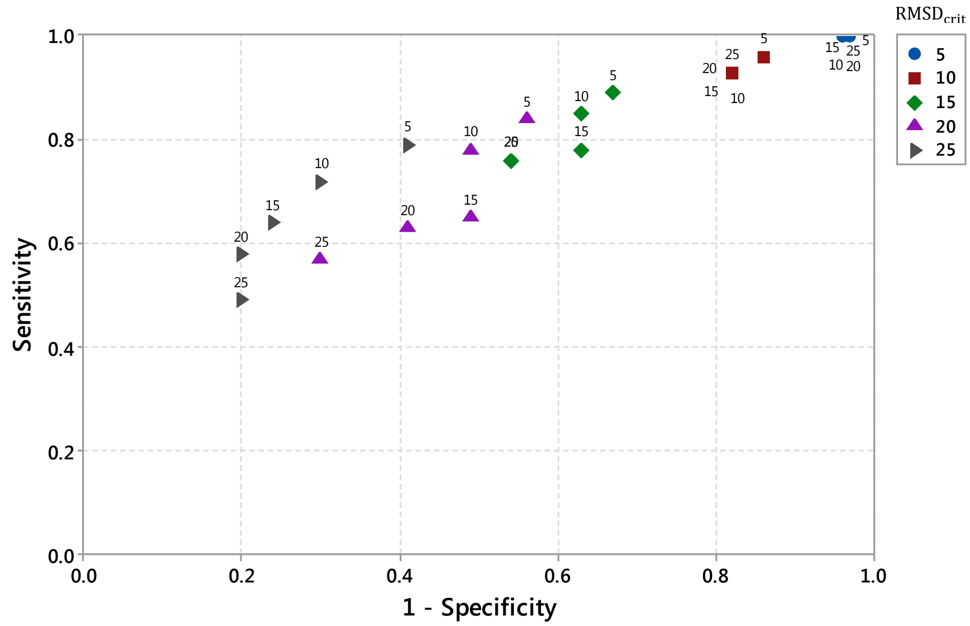


Figure A – 52 ROC curve: evaluation of parameters (RMSDcrit, ε) on participant 10's tracking task performance

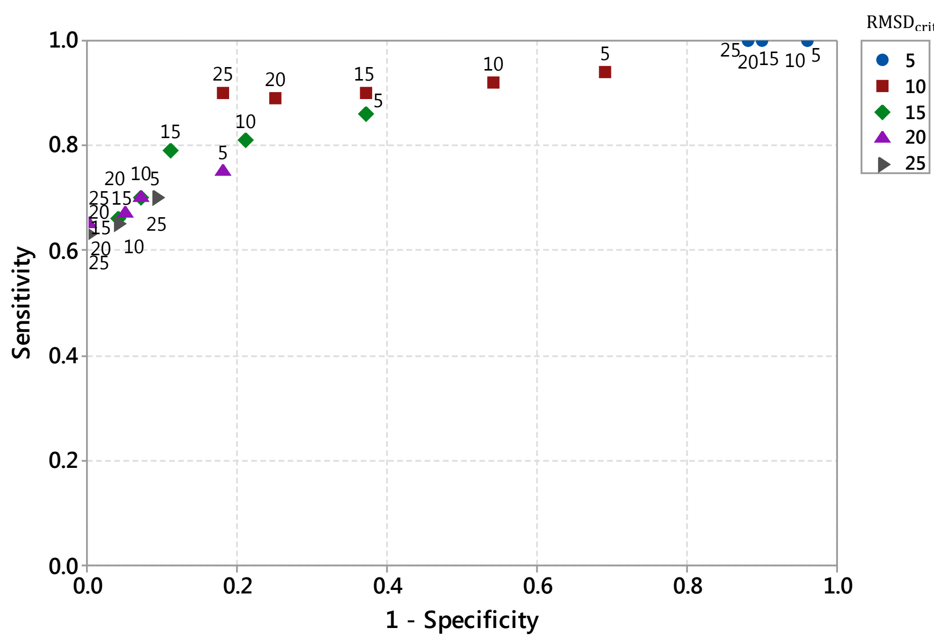


Figure A – 53 ROC curve: evaluation of parameters (RMSDcrit, ε) on participant 11's tracking task performance

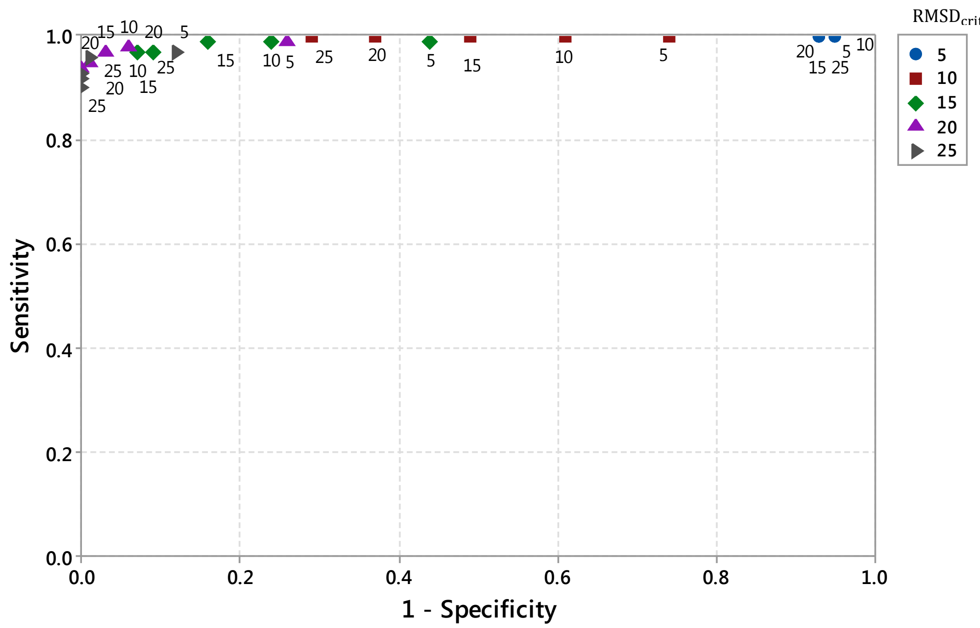


Figure A – 54 ROC curve: evaluation of parameters (RMSDcrit, ε) on participant 12's tracking task performance

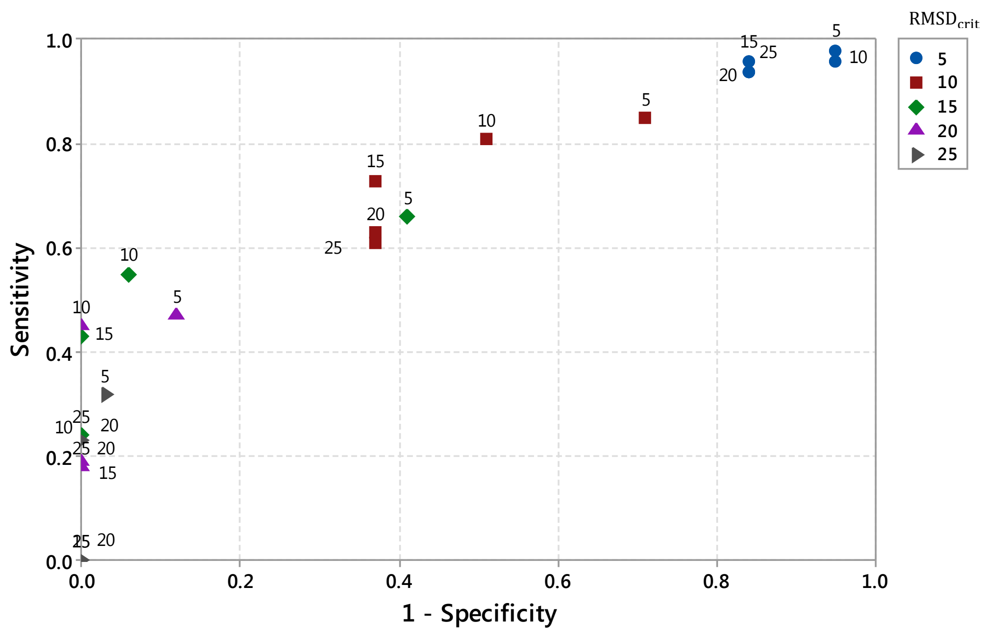


Figure A – 55 ROC curve: evaluation of parameters (RMSDcrit, ε) on participant 13's tracking task performance

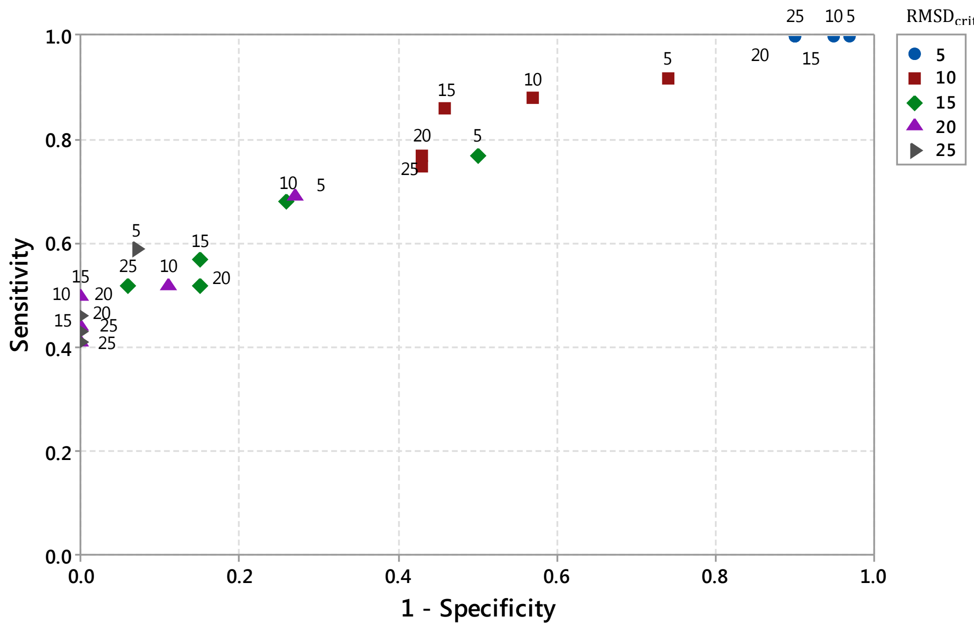


Figure A – 56 ROC curve: evaluation of parameters (RMSDcrit, ε) on participant 14's tracking task performance

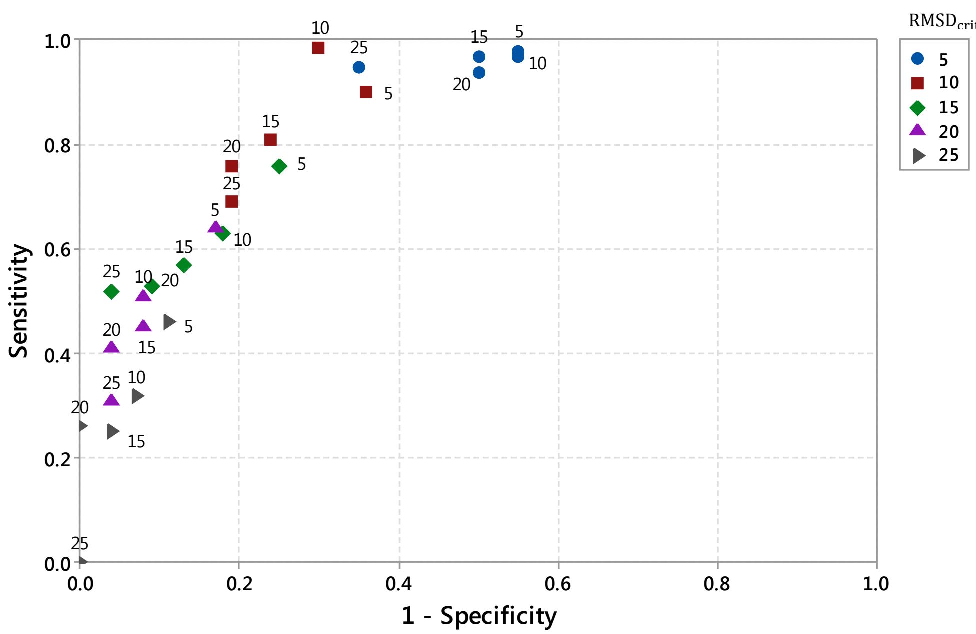


Figure A – 57 ROC curve: evaluation of parameters (RMSDcrit, ε) on participant 15's tracking task performance

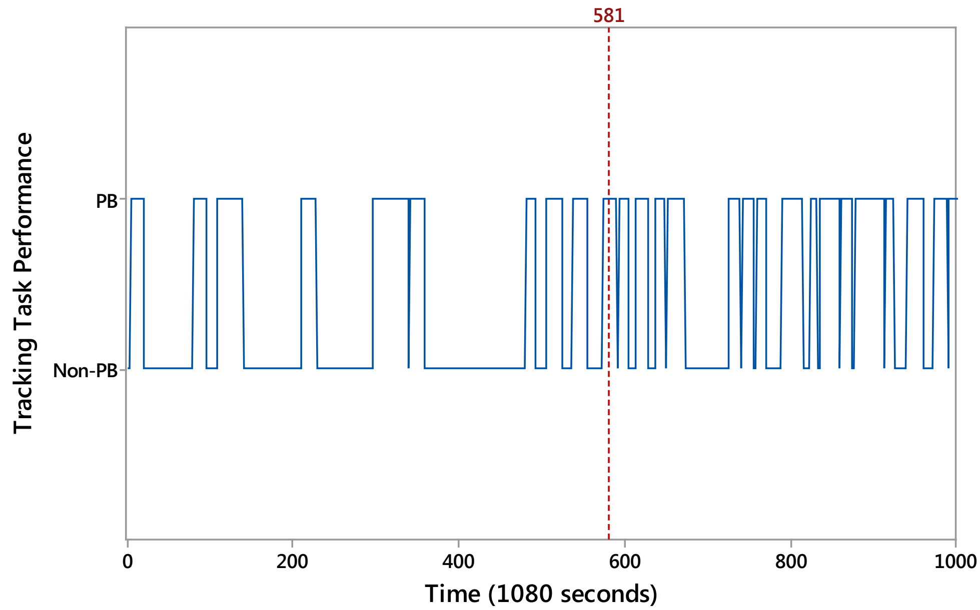


Figure A – 58 PB detection (RMSDcrit = 15, ε= 10) on participant 1's tracking performance data (criterion 1)

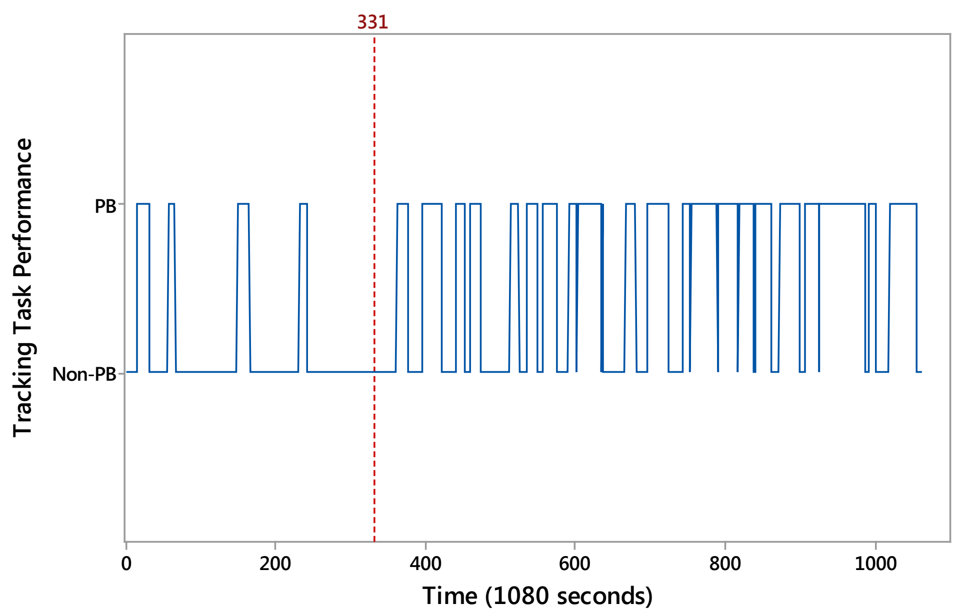


Figure A – 59 PB detection (RMSDcrit = 10, ε= 10) on participant 2's tracking performance data (criterion 1)

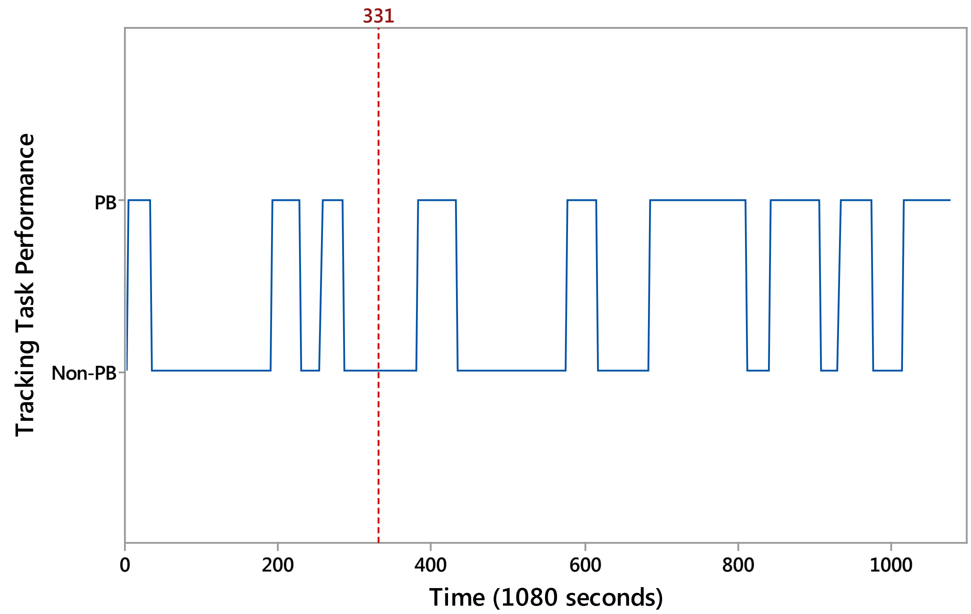


Figure A – 60 PB detection (RMSDcrit = 10, ε= 25) on participant 4's tracking performance data (criterion 1)

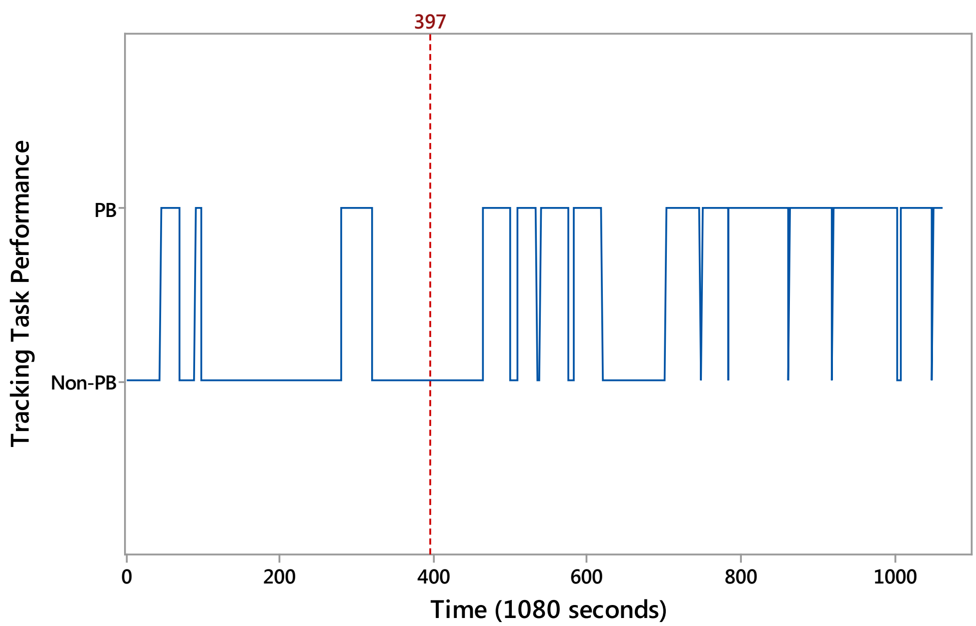


Figure A – 61 PB detection (RMSDcrit = 15, ε= 10) on participant 5's tracking performance data (criterion 1)

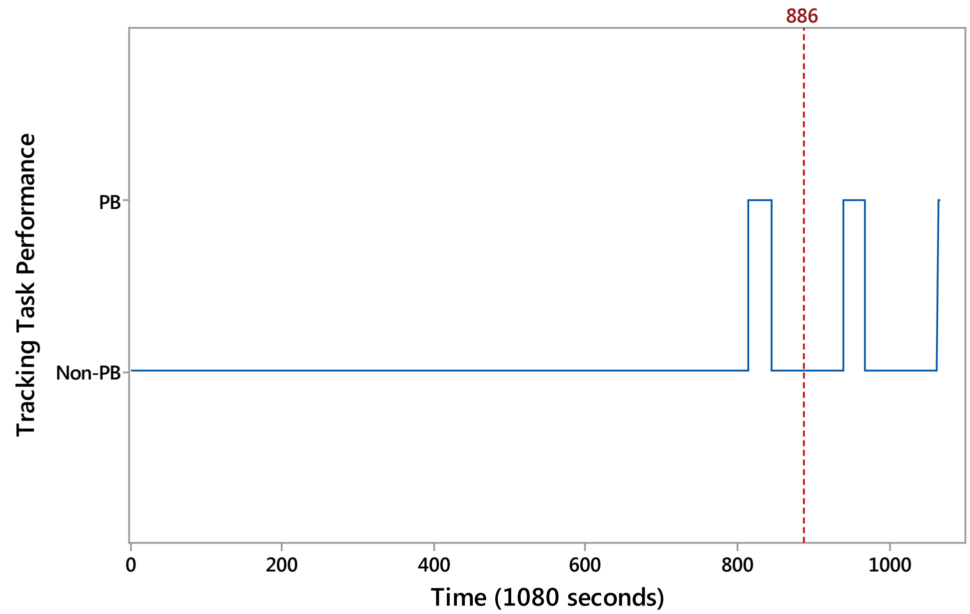


Figure A – 62 PB detection (RMSDcrit = 25, ε= 25) on participant 6's tracking performance data (criterion 1)

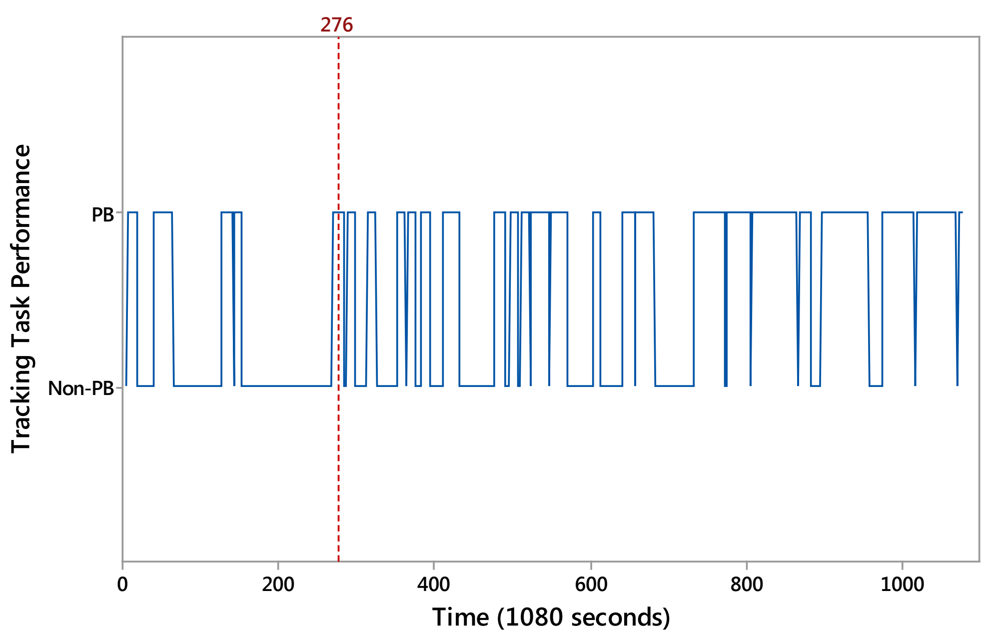


Figure A – 63 PB detection (RMSDcrit = 15, ε= 10) on participant 7's tracking performance data (criterion 1)

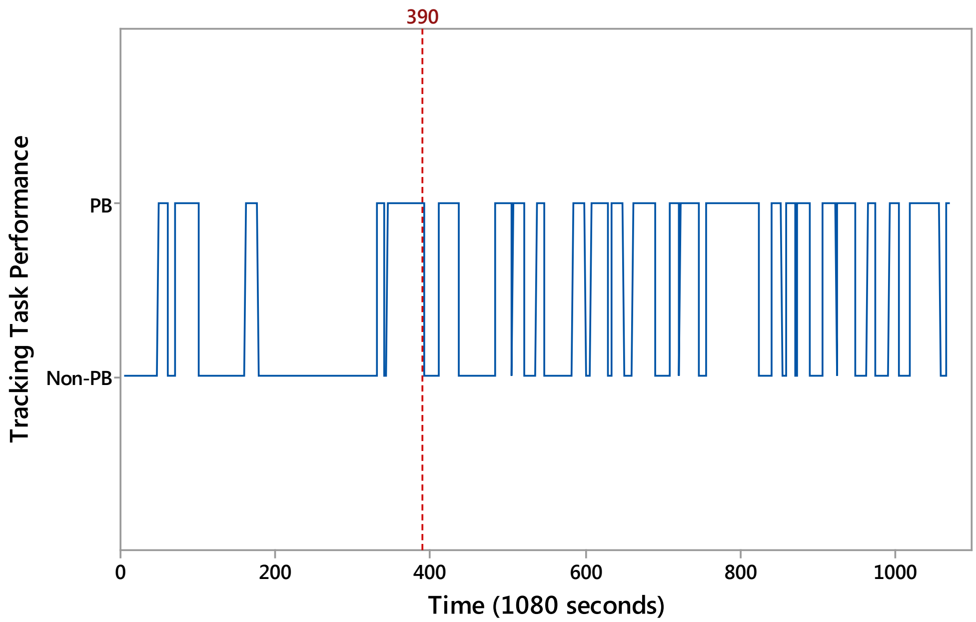


Figure A – 64 PB detection (RMSDcrit = 25, ε= 10) on participant 10's tracking performance data (criterion 1)

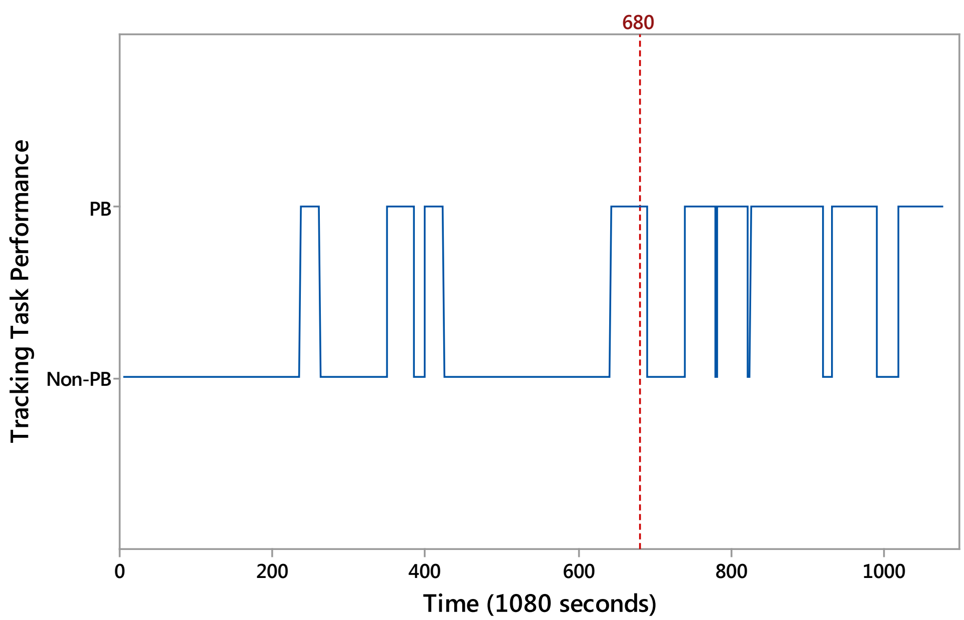


Figure A – 65 PB detection (RMSDcrit t = 10, ε= 25) on participant 11's tracking performance data (criterion 1)

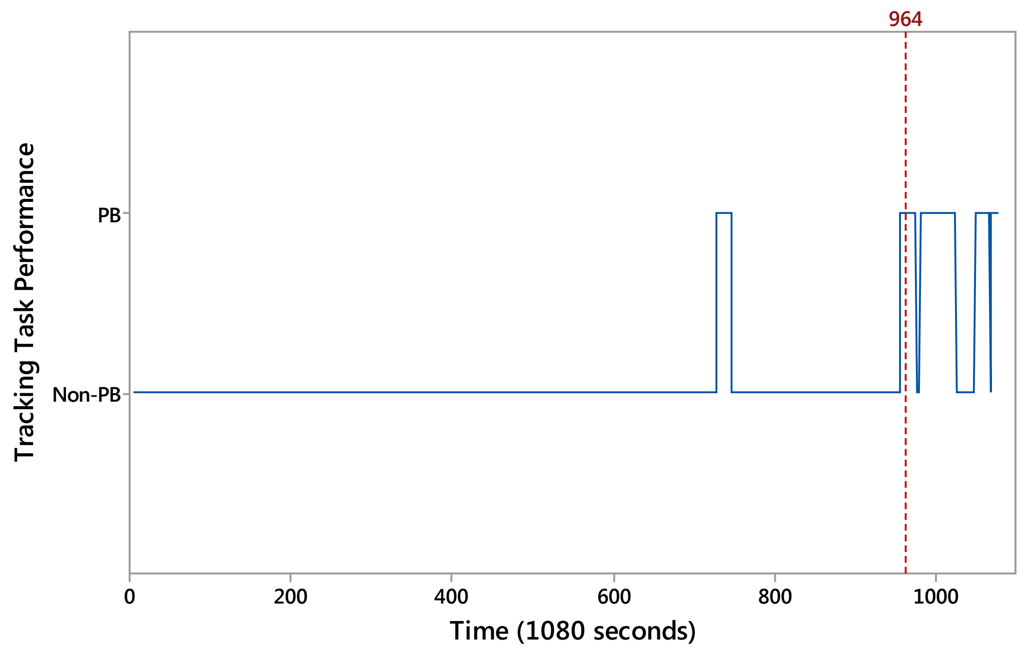


Figure A – 66 PB detection (RMSDcrit = 20, ε= 15) on participant 12's tracking performance data (criterion 1)

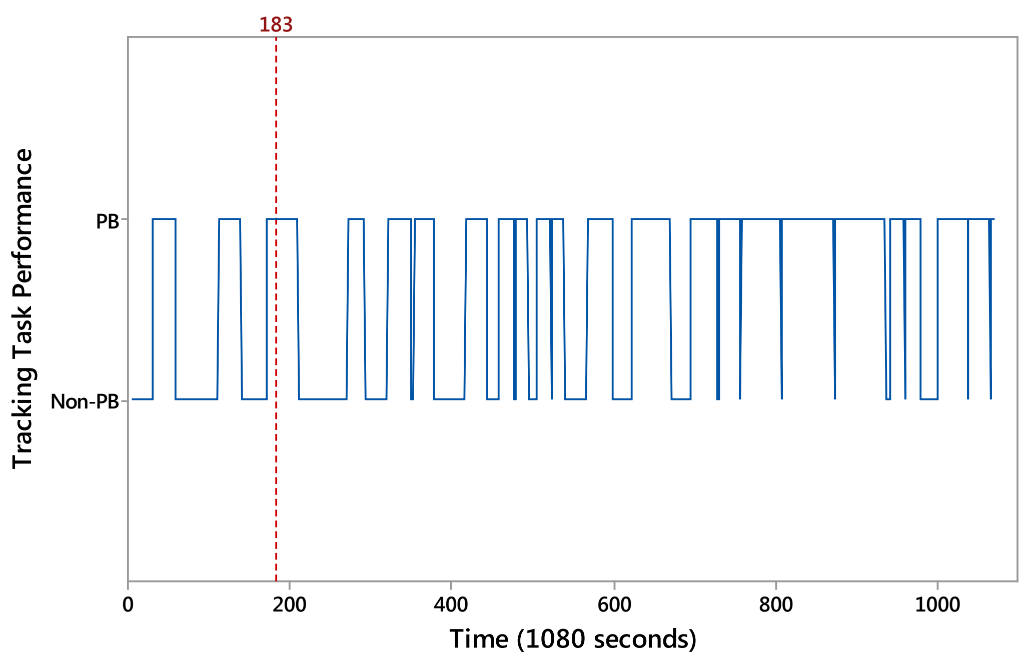


Figure A – 67 PB detection (RMSDcrit = 10, ε= 15) on participant 13's tracking performance data (criterion 1)

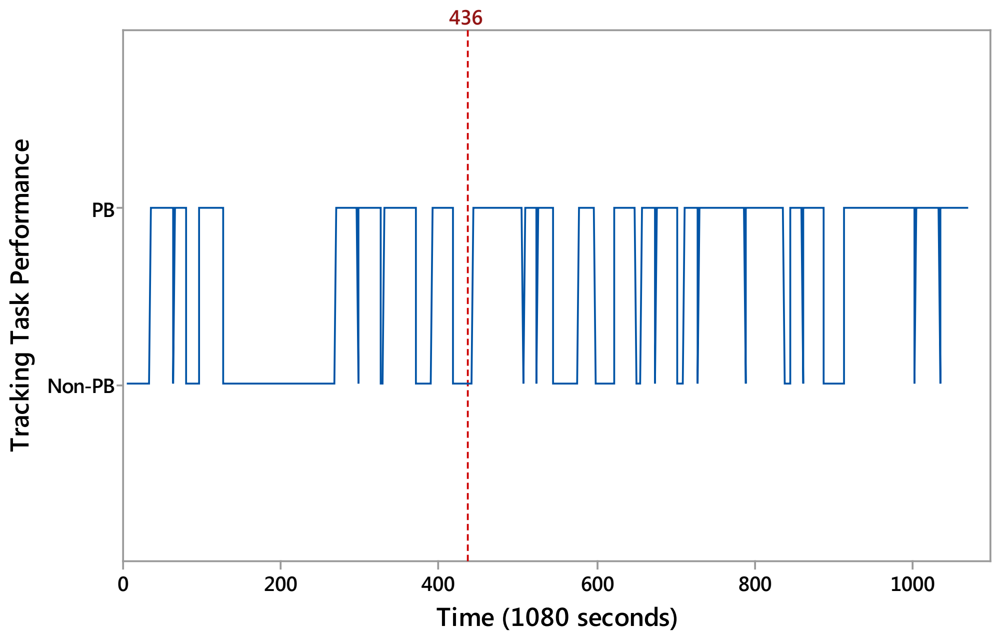


Figure A – 68 PB detection (RMSDcrit = 20, ε= 5) on participant 14’s tracking performance data (criterion 1)

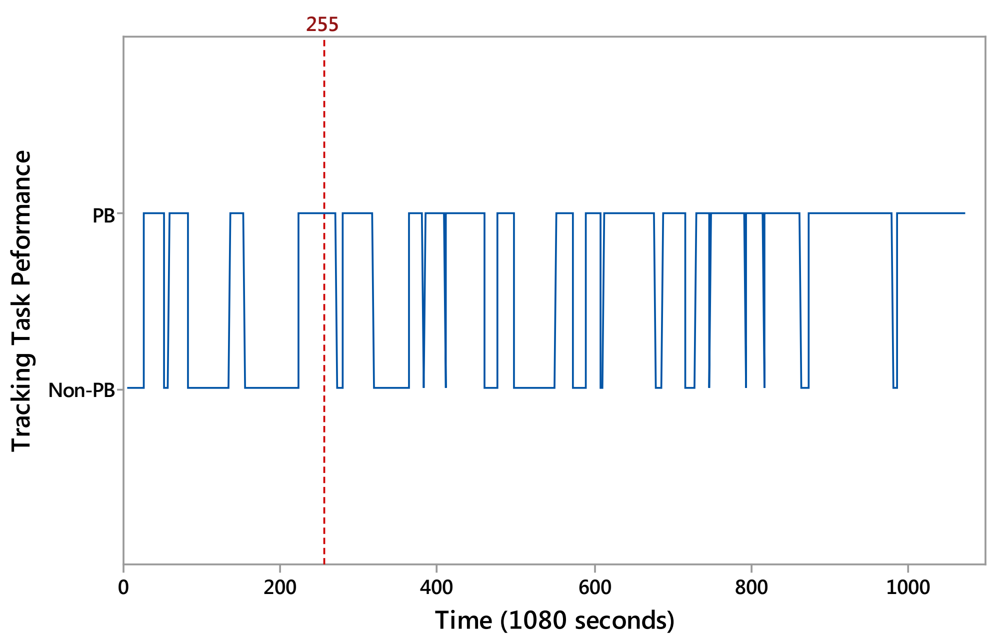


Figure A – 69 PB detection (RMSDcrit = 10, ε= 15) on participant 15’s tracking performance data (criteria 1)

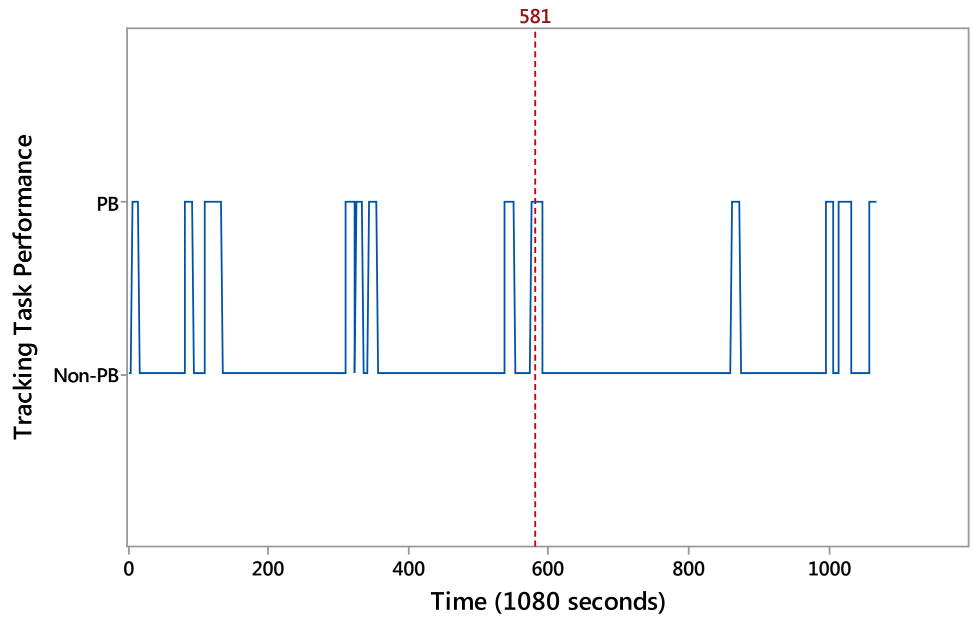


Figure A – 70 PB detection (RMSDcrit = 20, ε= 10) on participant 1’s tracking performance data (criterion 2)

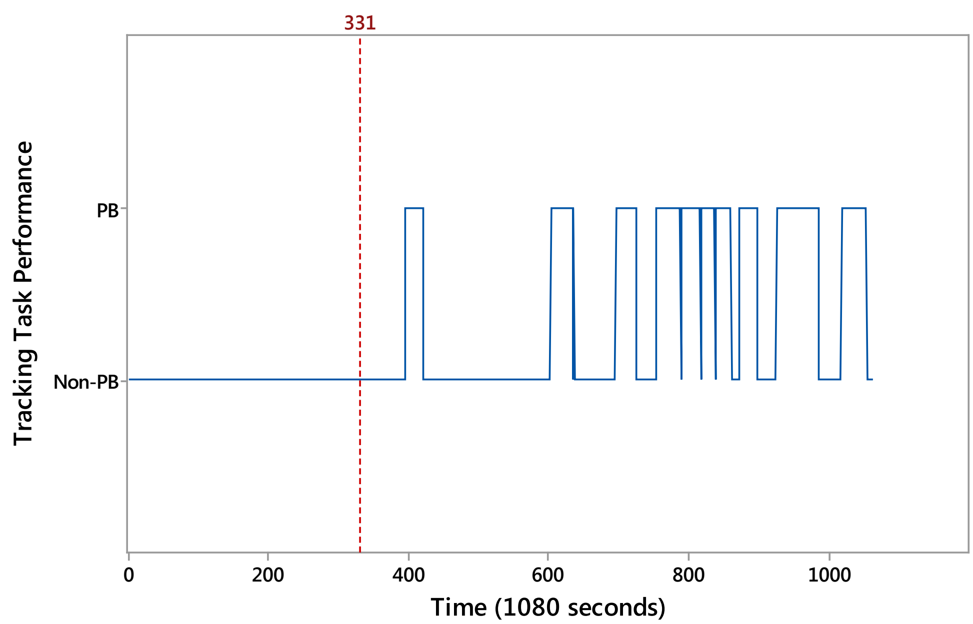


Figure A – 71 PB detection (RMSDcrit = 10, ε= 20) on participant 2’s tracking performance data (criterion 2)

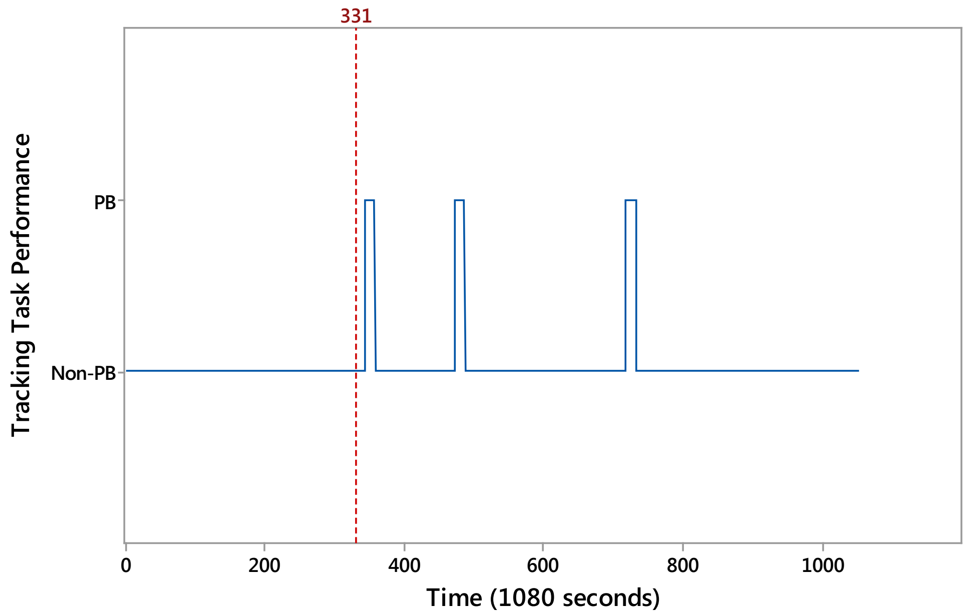


Figure A – 72 PB detection (RMSDcrit = 20, ε= 15) on participant 4's tracking performance data (criterion 2)

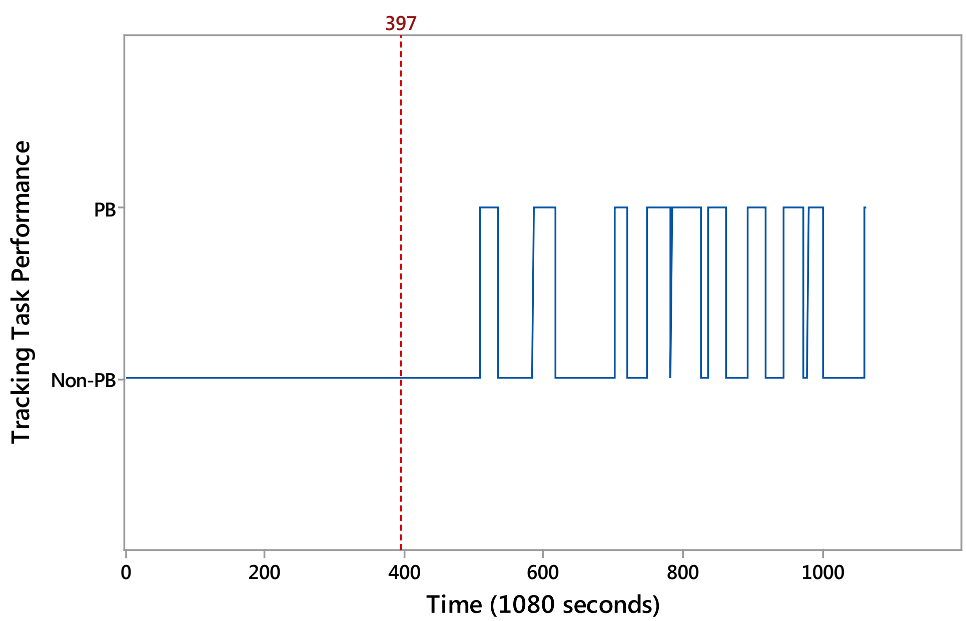


Figure A – 73 PB detection (RMSDcrit = 15, ε= 20) on participant 5's tracking performance data (criterion 2)

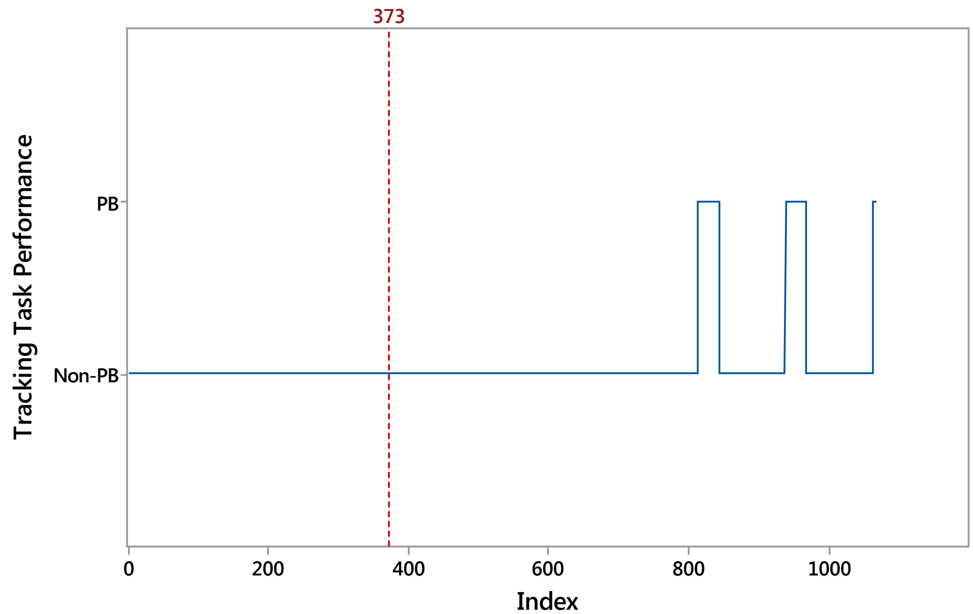


Figure A – 74 PB detection (RMSDcrit = 20, ε= 15) on participant 6's tracking performance data (criterion 2)

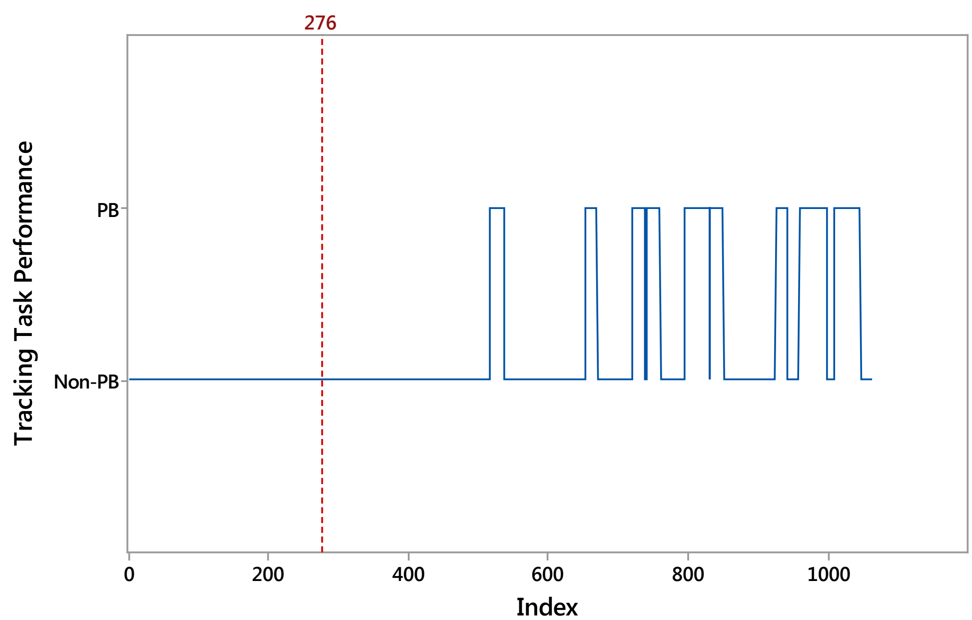


Figure A – 75 PB detection (RMSDcrit = 25, ε= 20) on participant 7's tracking performance data (criterion 2)

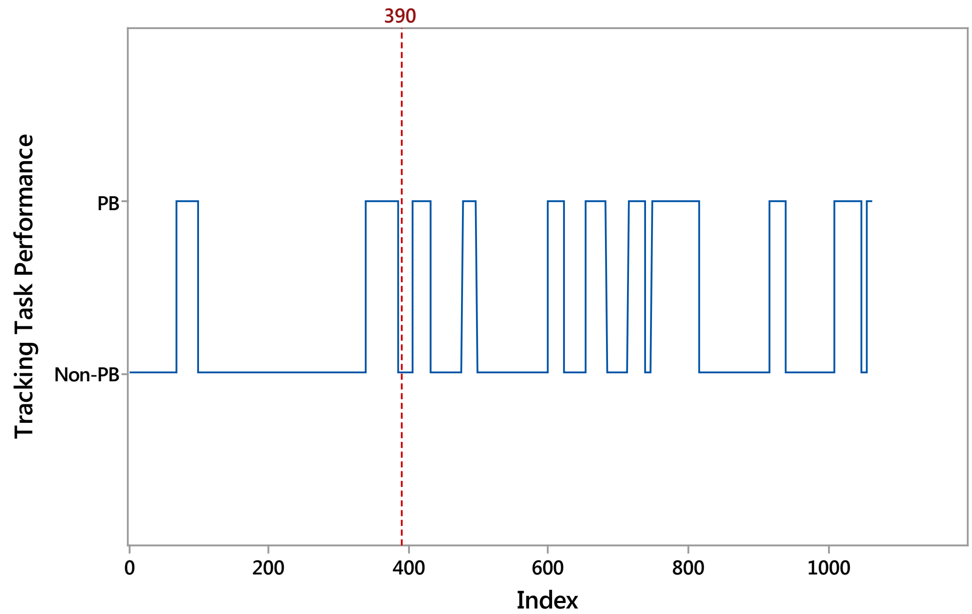


Figure A – 76 PB detection (RMSDcrit = 20, ε= 15) on participant 10's tracking performance data (criterion 2)

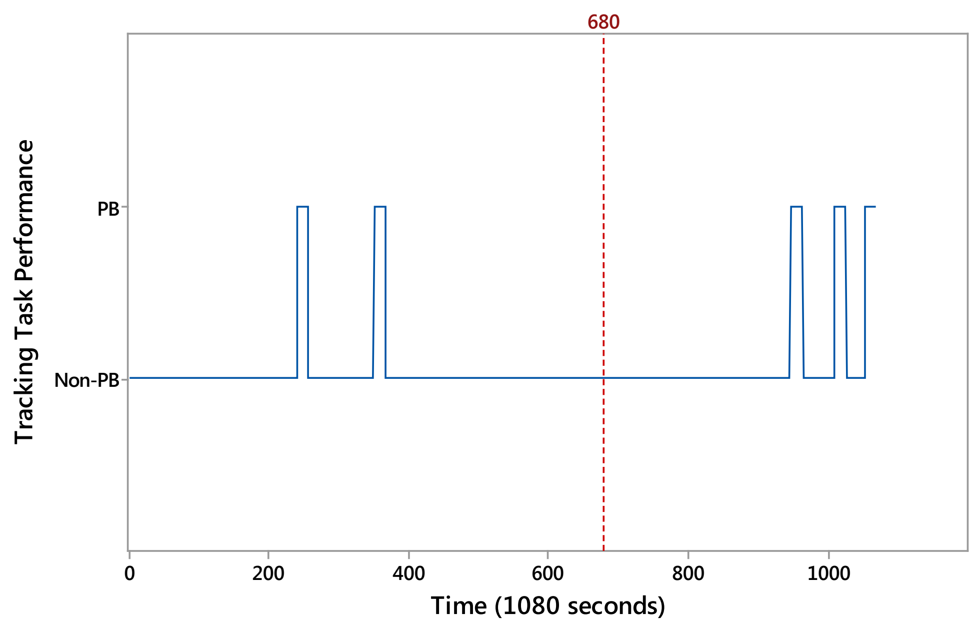


Figure A – 77 PB detection (RMSDcrit = 25, ε= 15) on participant 11's tracking performance data (criterion 2)

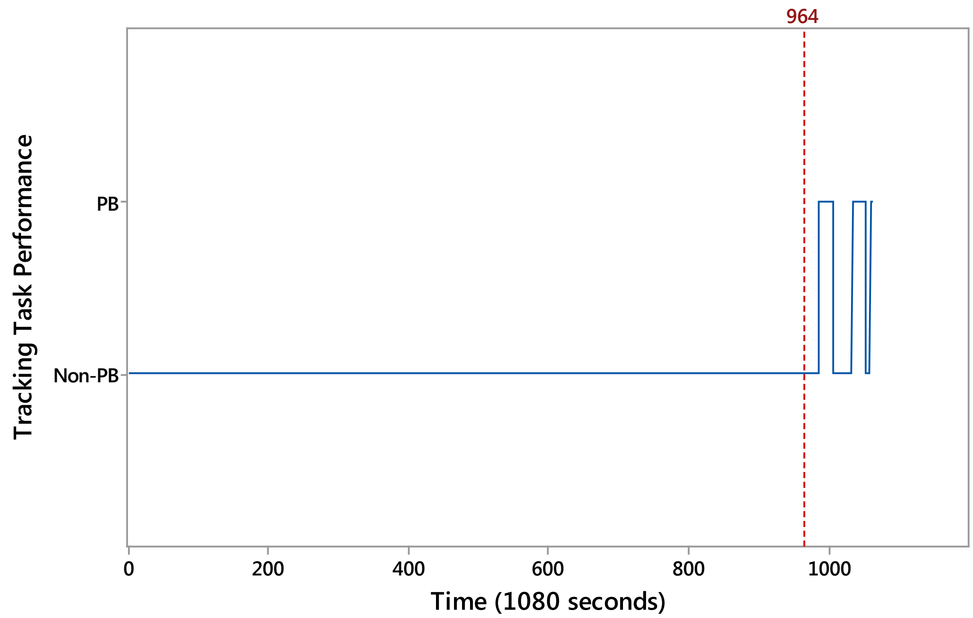


Figure A – 78 PB detection (RMSDcrit = 25, ε= 15) on participant 12's tracking performance data (criterion 2)

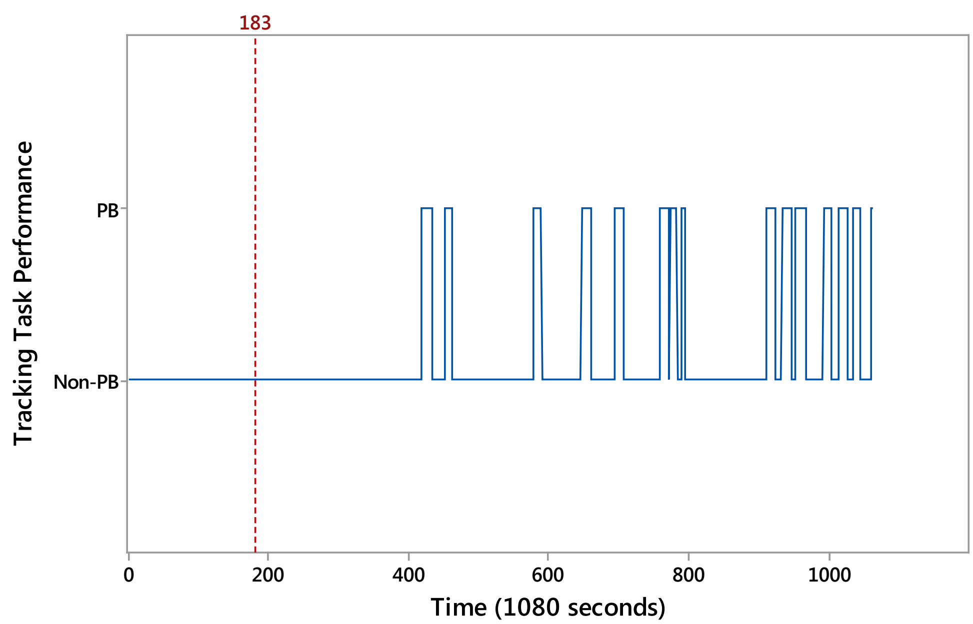


Figure A – 79 PB detection (RMSDcrit = 20, ε= 10) on participant 13's tracking performance data (criterion 2)

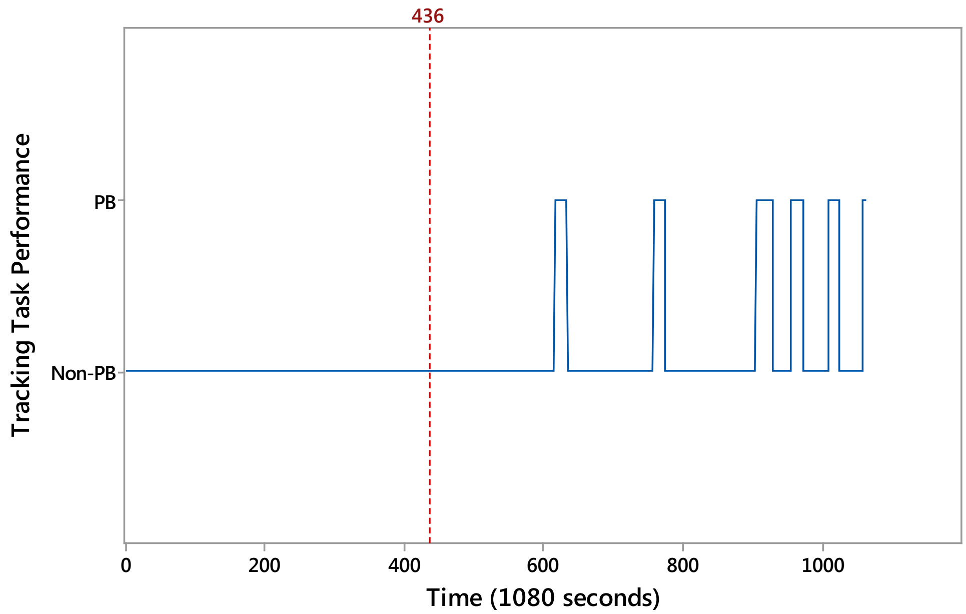


Figure A – 80 PB detection (RMSDcrit = 20, ε= 15) on participant 14's tracking performance data (criterion 2)

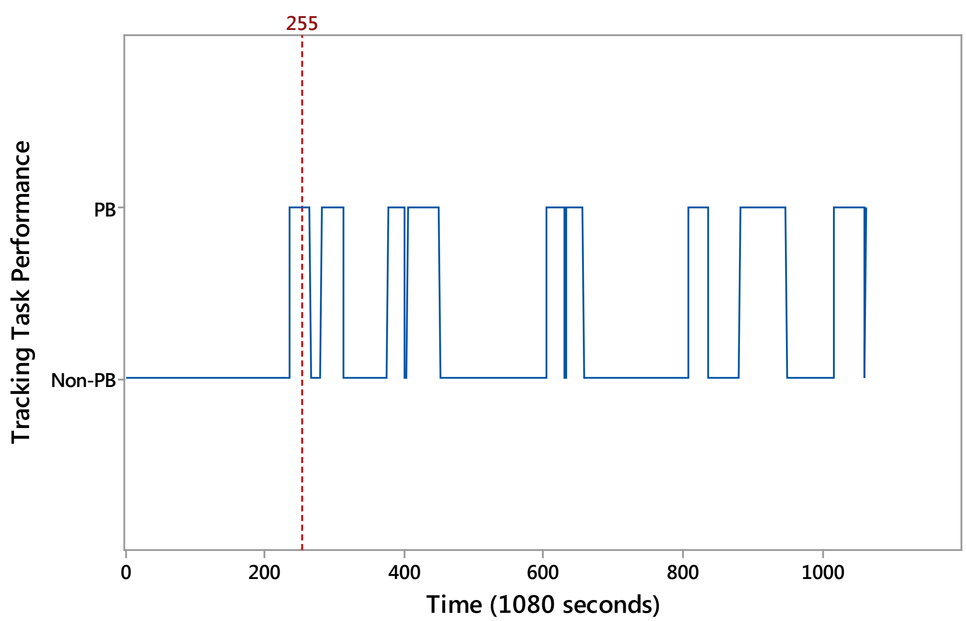


Figure A – 81 PB detection (RMSDcrit = 15, ε= 25) on participant 15's tracking performance data (criterion 2)

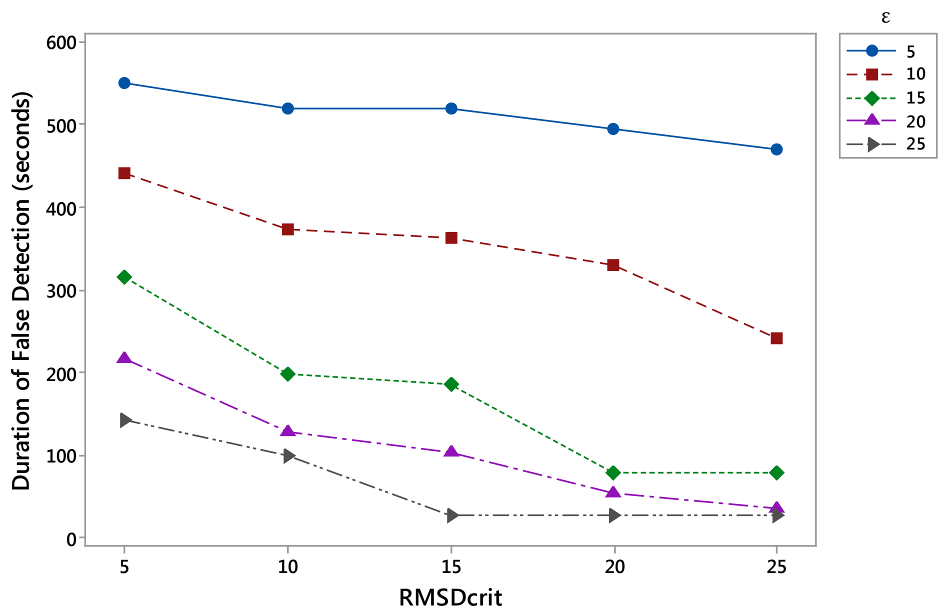


Figure A – 81 Participant 1’s duration of false detection (seconds) vs. $RMSD_{crit}$= (5, 10, 15, 20, and 25)

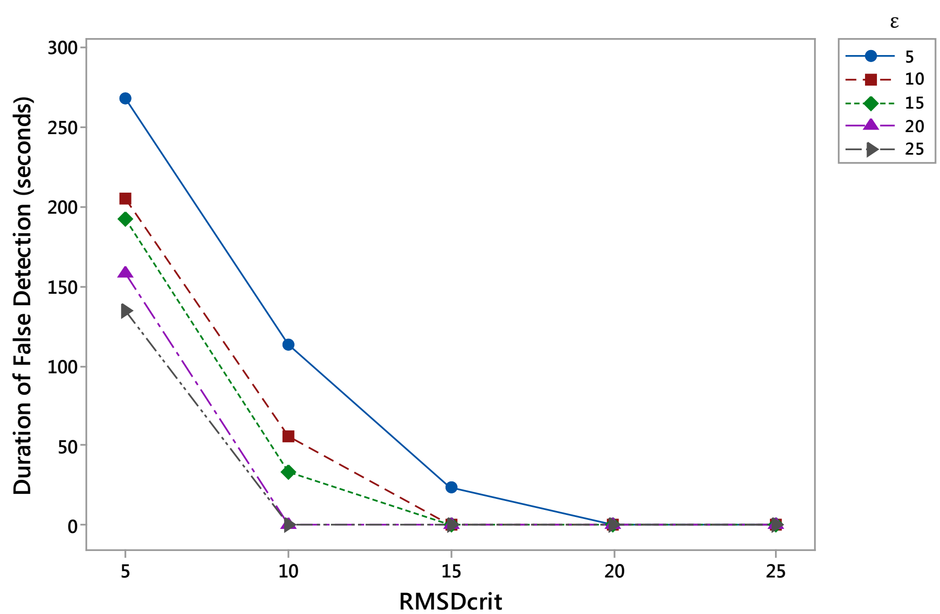


Figure A – 82 Participant 2’s duration of false detection (seconds) vs. $RMSD_{crit}$= (5, 10, 15, 20, and 25)

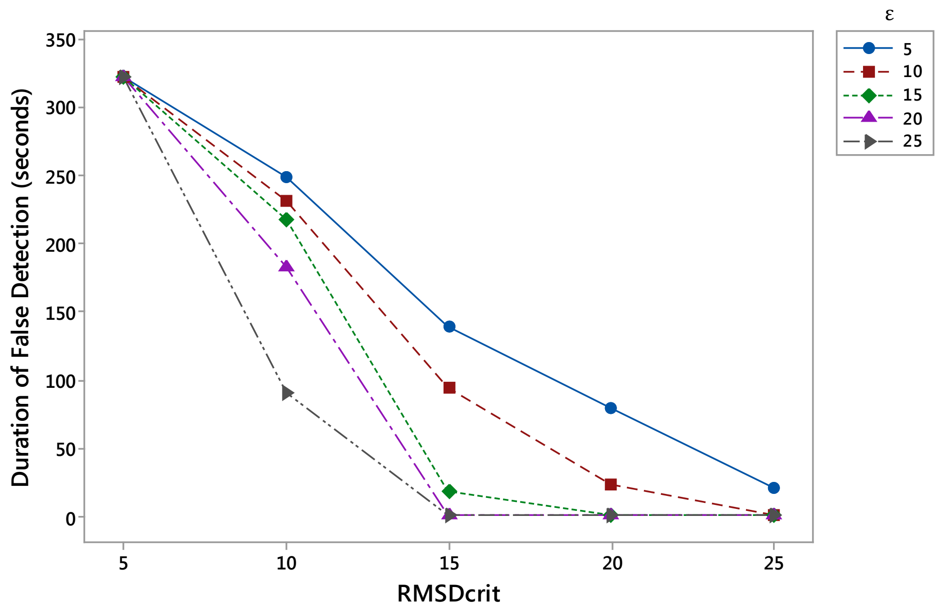


Figure A – 83 Participant 4's duration of false detection (seconds) vs.
$RMSD_{crit}$= (5, 10, 15, 20, and 25)

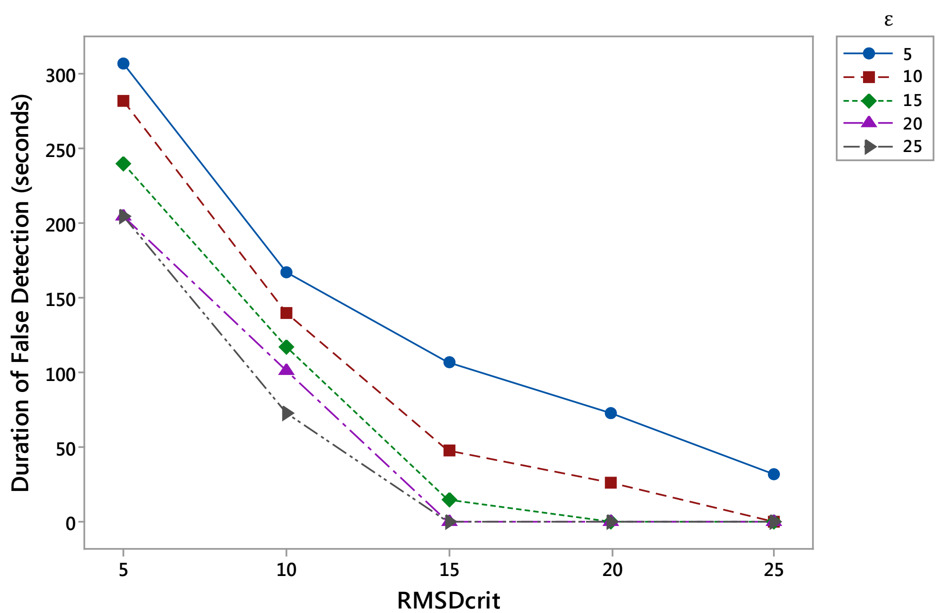


Figure A – 84 Participant 5's duration of false detection (seconds) vs.
$RMSD_{crit}$= (5, 10, 15, 20, and 25)

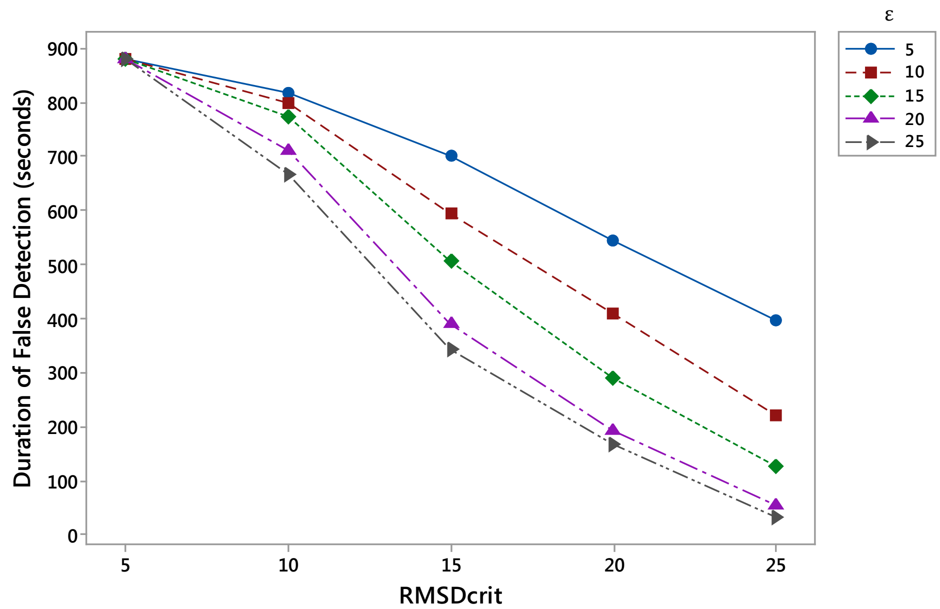


Figure A – 85 Participant 6's duration of false detection (seconds) vs. $RMSD_{crit}$= (5, 10, 15, 20, and 25)

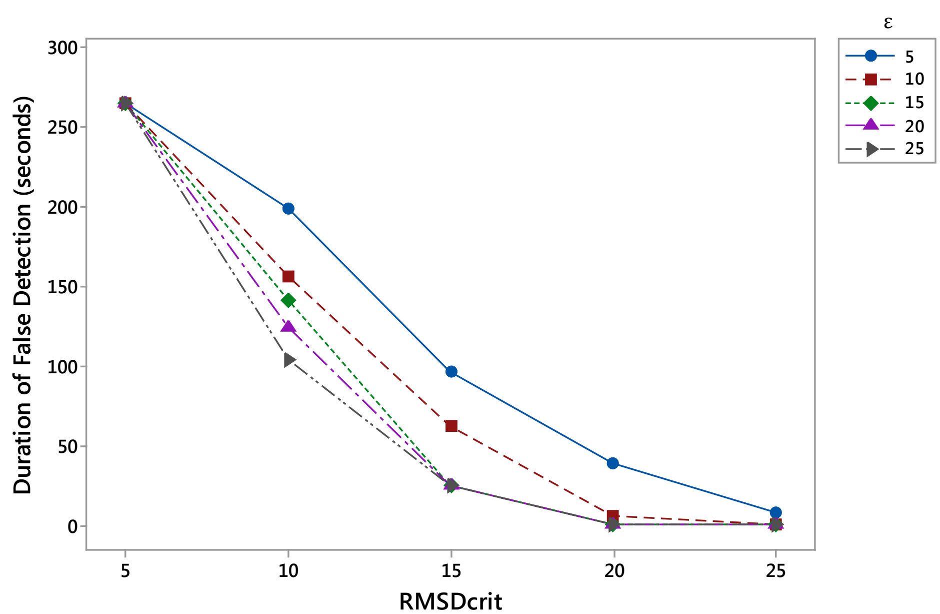


Figure A – 86 Participant 7's duration of false detection (seconds) vs. $RMSD_{crit}$= (5, 10, 15, 20, and 25)

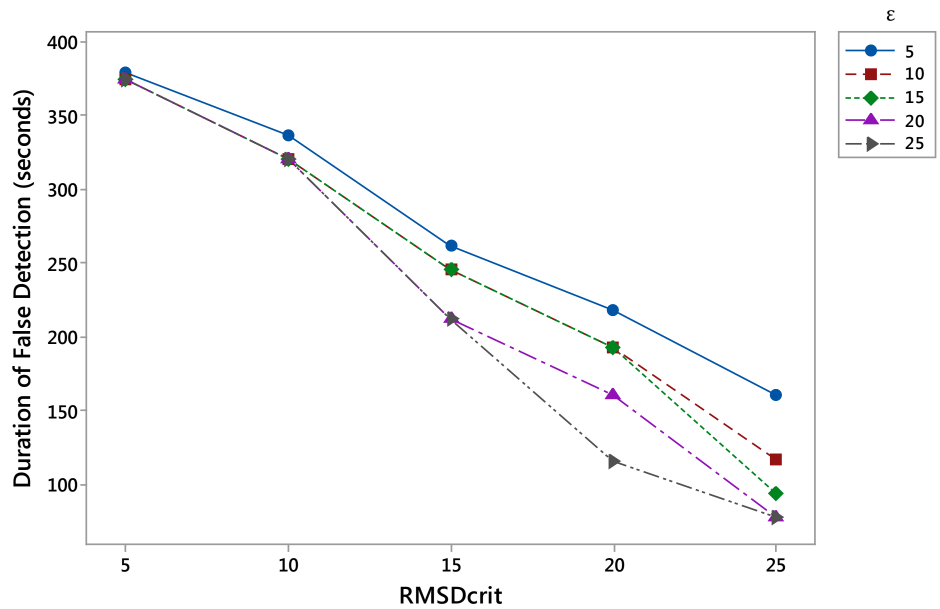


Figure A – 87 Participant 10's duration of false detection (seconds) vs. $RMSD_{crit}$= (5, 10, 15, 20, and 25)

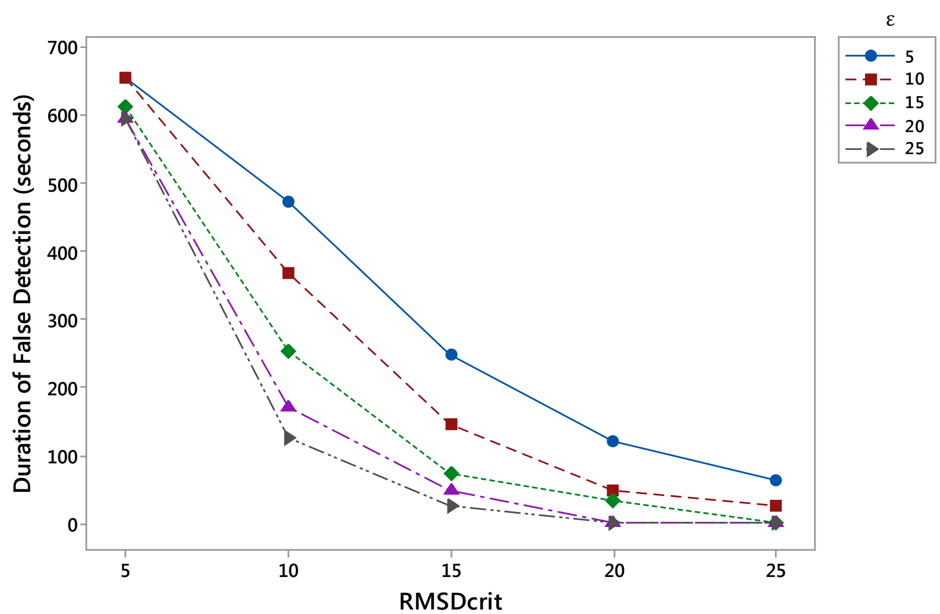


Figure A – 88 Participant 11's duration of false detection (seconds) vs. $RMSD_{crit}$= (5, 10, 15, 20, and 25)

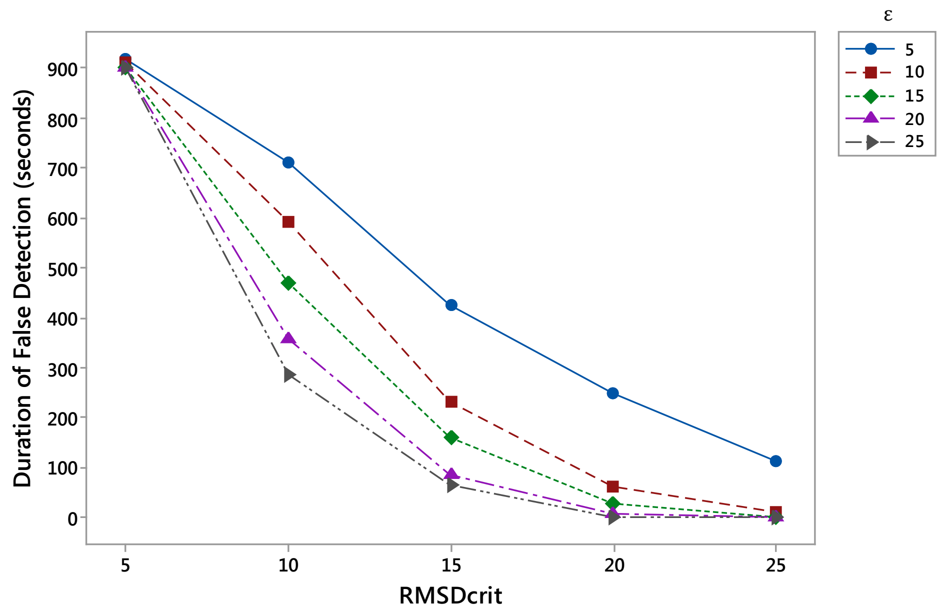


Figure A – 89 Participant 12's duration of false detection (seconds) vs. $RMSD_{crit}$= (5, 10, 15, 20, and 25)

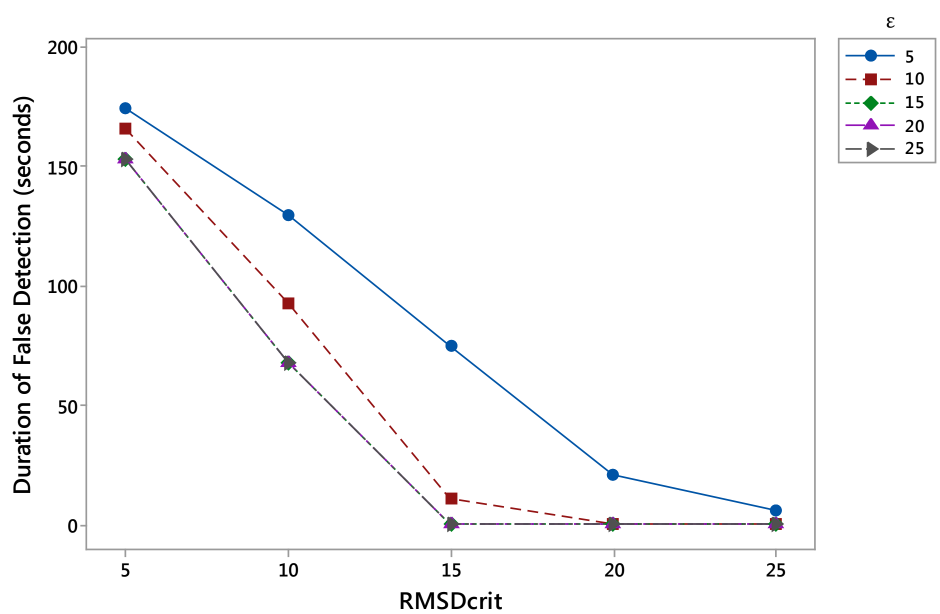


Figure A – 89 Participant 13's duration of false detection (seconds) vs. $RMSD_{crit}$= (5, 10, 15, 20, and 25)

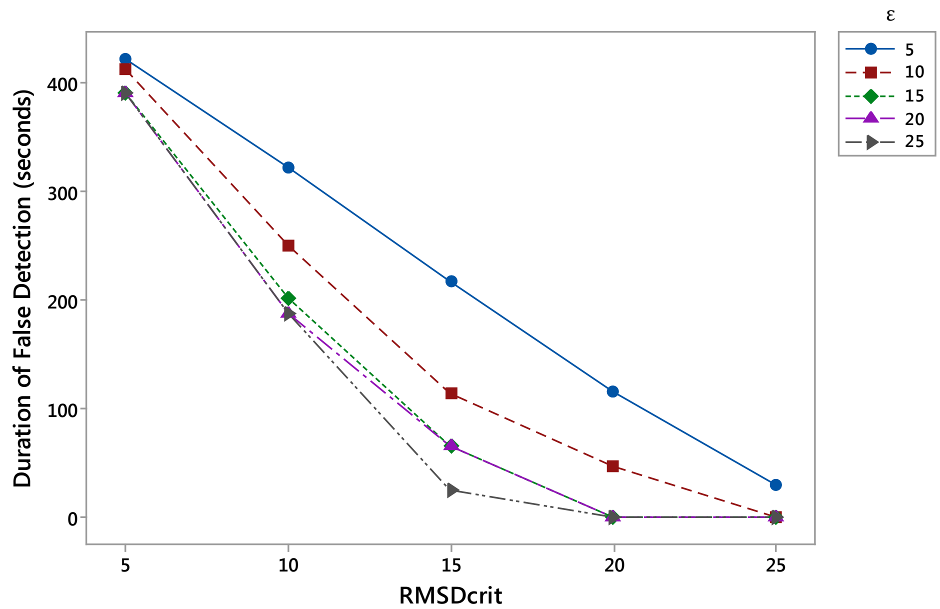


Figure A – 90 Participant 14's duration of false detection (seconds) vs. $RMSD_{crit}$= (5, 10, 15, 20, and 25)

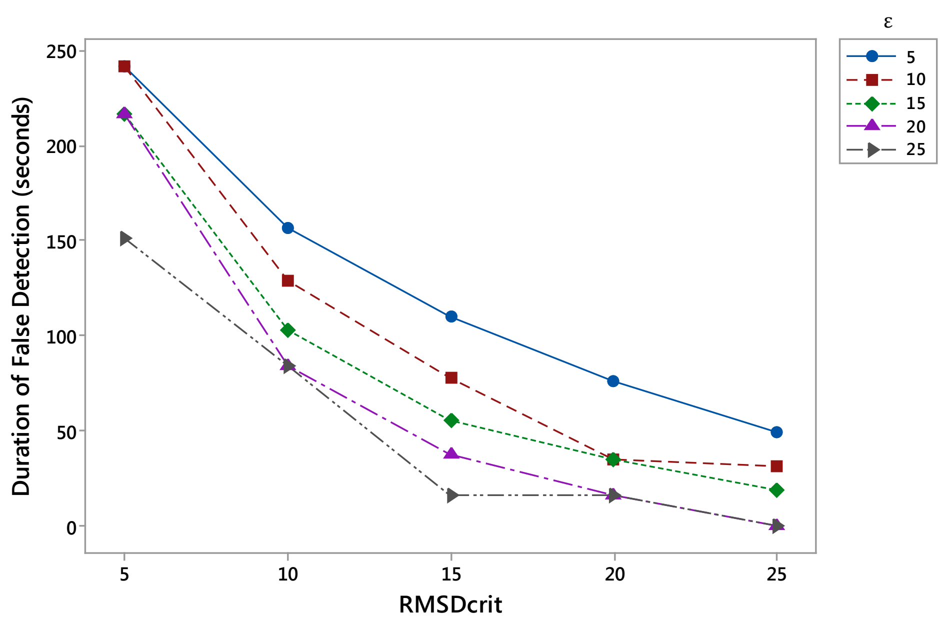


Figure A – 91 Participant 15's duration of false detection (seconds) vs. $RMSD_{crit}$= (5, 10, 15, 20, and 25)

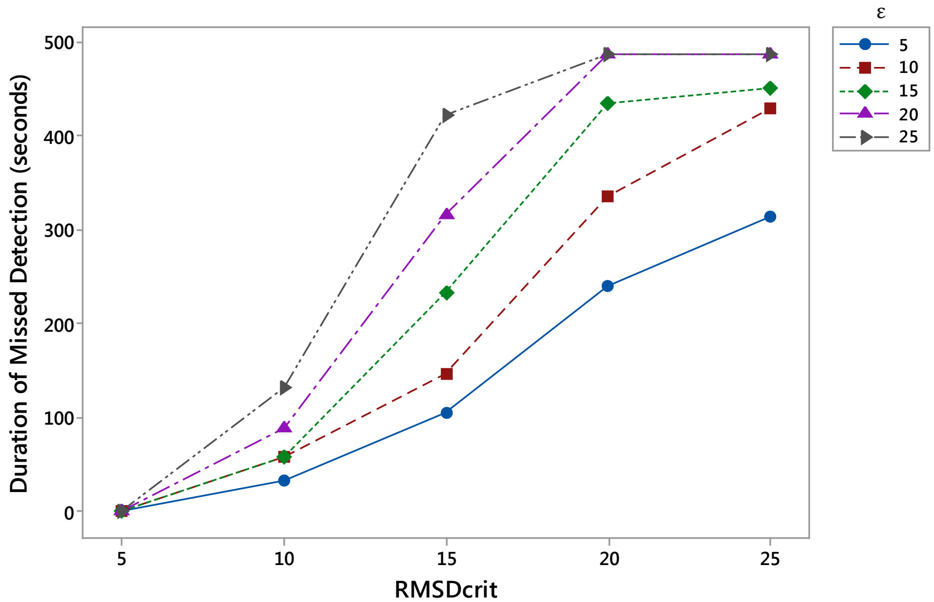


Figure A – 92 Participant 1’s duration of missed detection (seconds) vs. $RMSD_{crit}$= (5, 10, 15, 20, and 25)

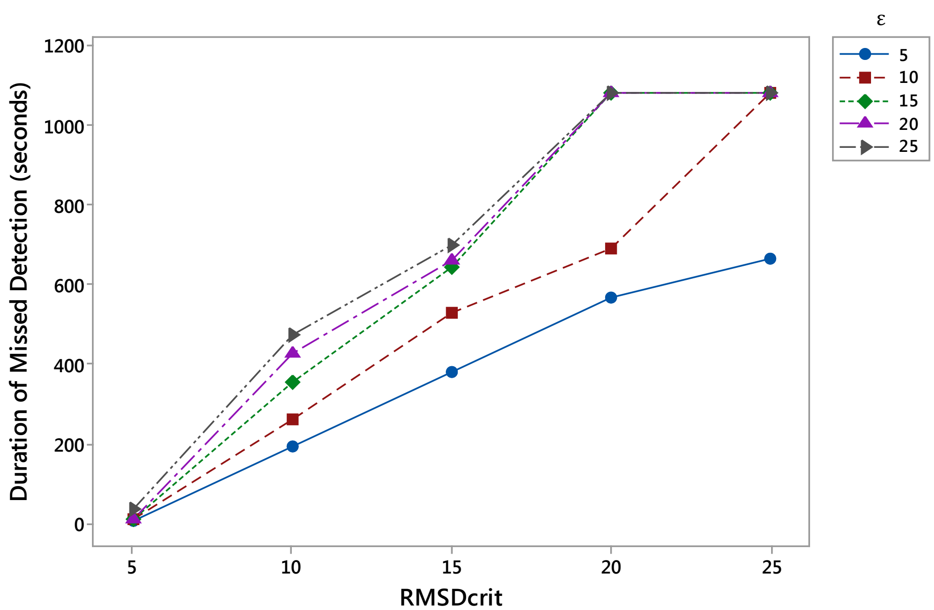


Figure A – 93 Participant 2’s duration of missed detection (seconds) vs. $RMSD_{crit}$= (5, 10, 15, 20, and 25)

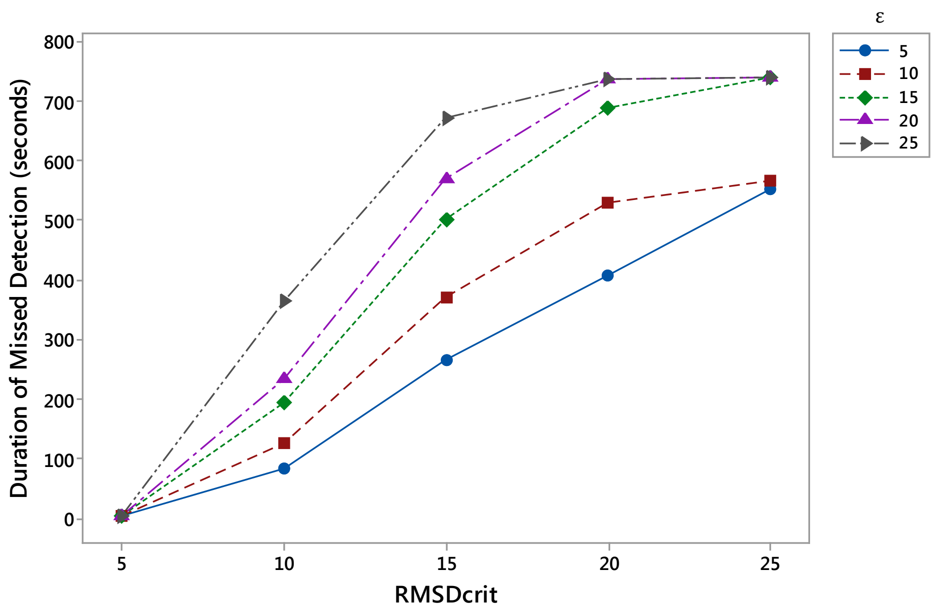


Figure A – 94 Participant 4's duration of missed detection (seconds) vs. $RMSD_{crit}$= (5, 10, 15, 20, and 25)

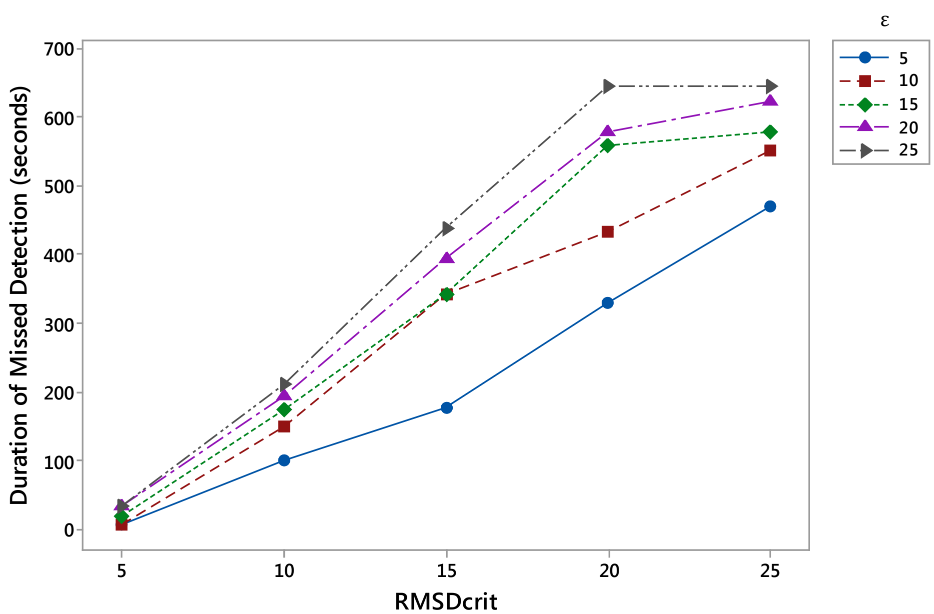


Figure A – 95 Participant 5's duration of missed detection (seconds) vs. $RMSD_{crit}$= (5, 10, 15, 20, and 25)

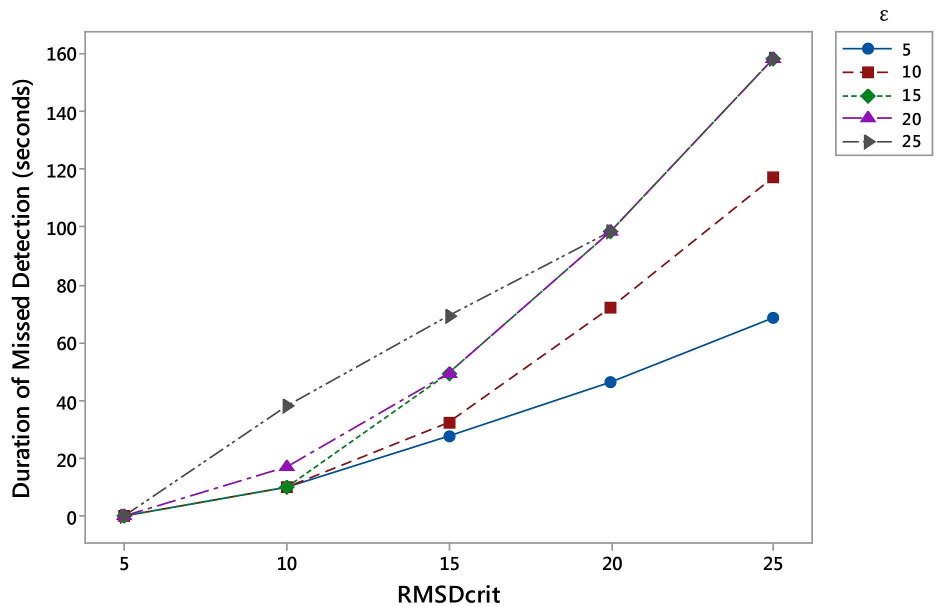


Figure A – 96 Participant 6's duration of missed detection (seconds) vs. $RMSD_{crit}$= (5, 10, 15, 20, and 25)

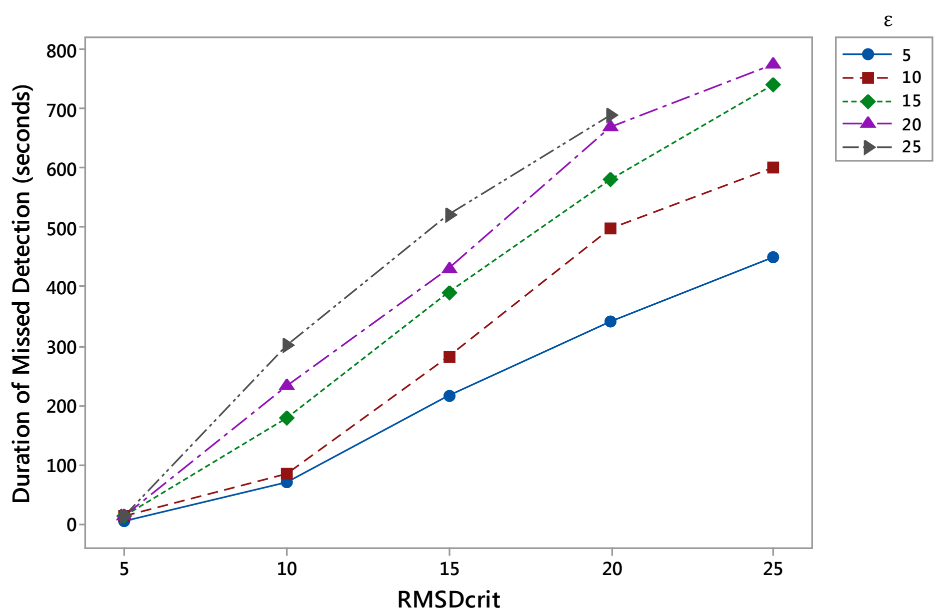


Figure A – 97 Participant 7's duration of missed detection (seconds) vs. $RMSD_{crit}$= (5, 10, 15, 20, and 25)

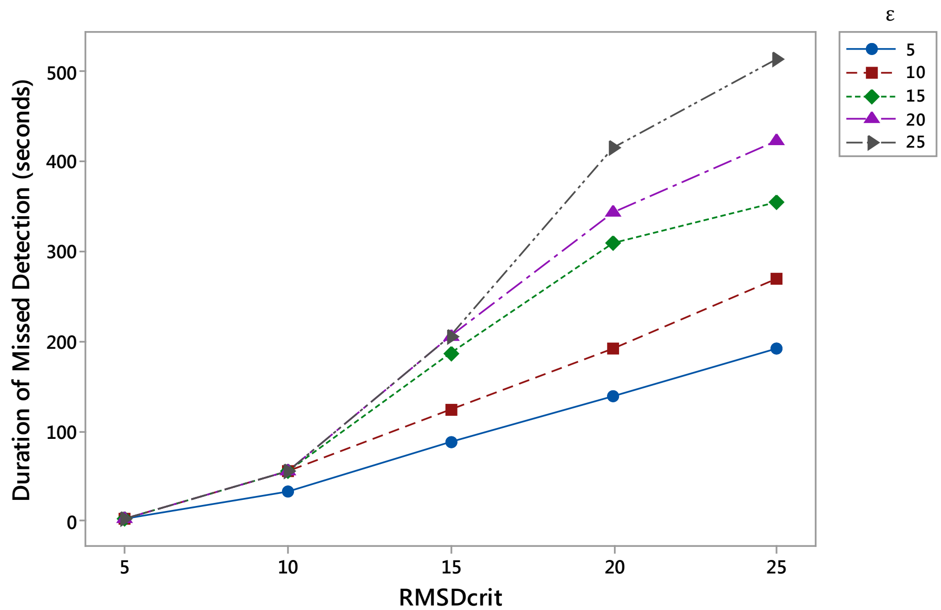


Figure A – 98 Participant 10's duration of missed detection (seconds) vs. $RMSD_{crit}$= (5, 10, 15, 20, and 25)

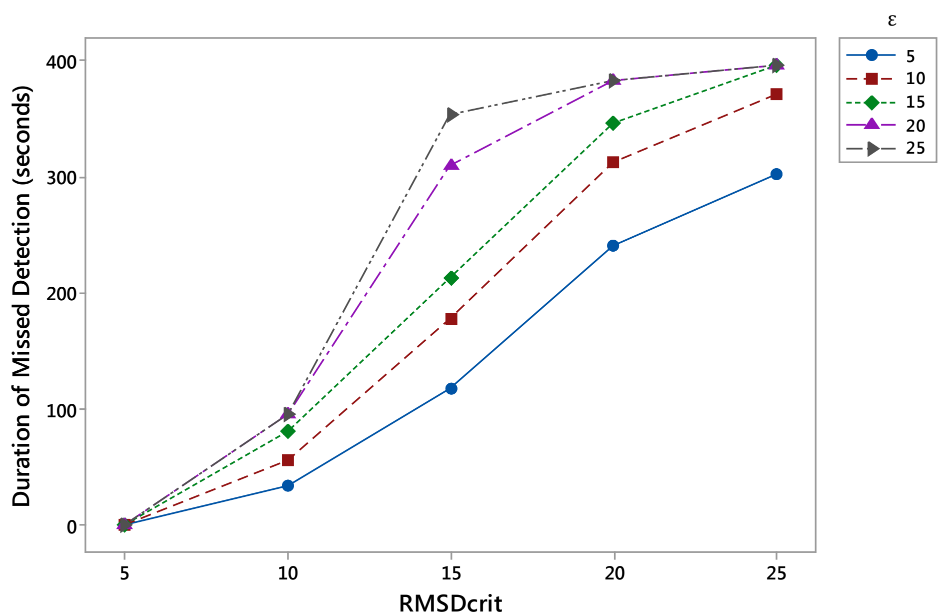


Figure A – 99 Participant 11's duration of missed detection (seconds) vs. $RMSD_{crit}$= (5, 10, 15, 20, and 25)

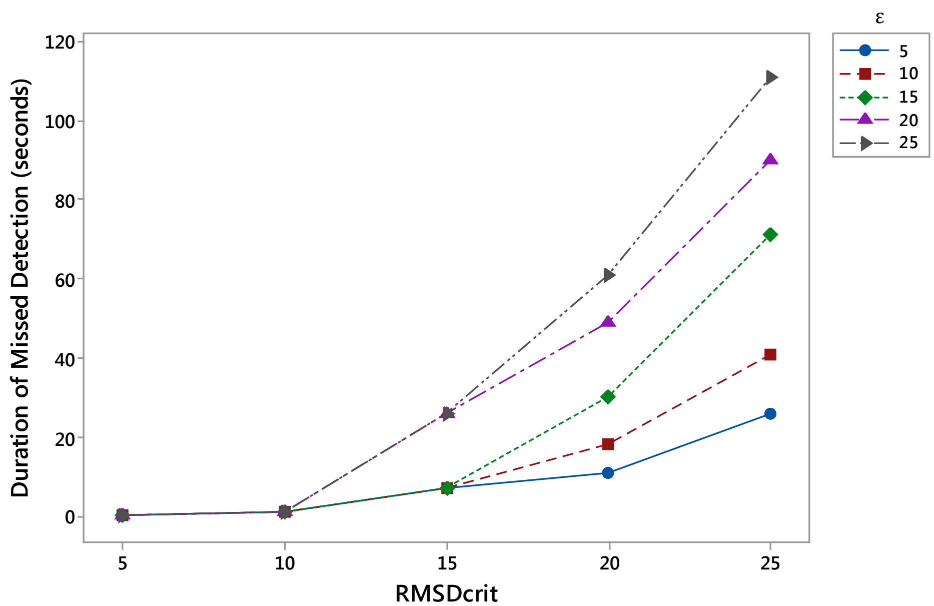


Figure A – 100 Participant 12’s duration of missed detection (seconds) vs. $RMSD_{crit}$= (5, 10, 15, 20, and 25)

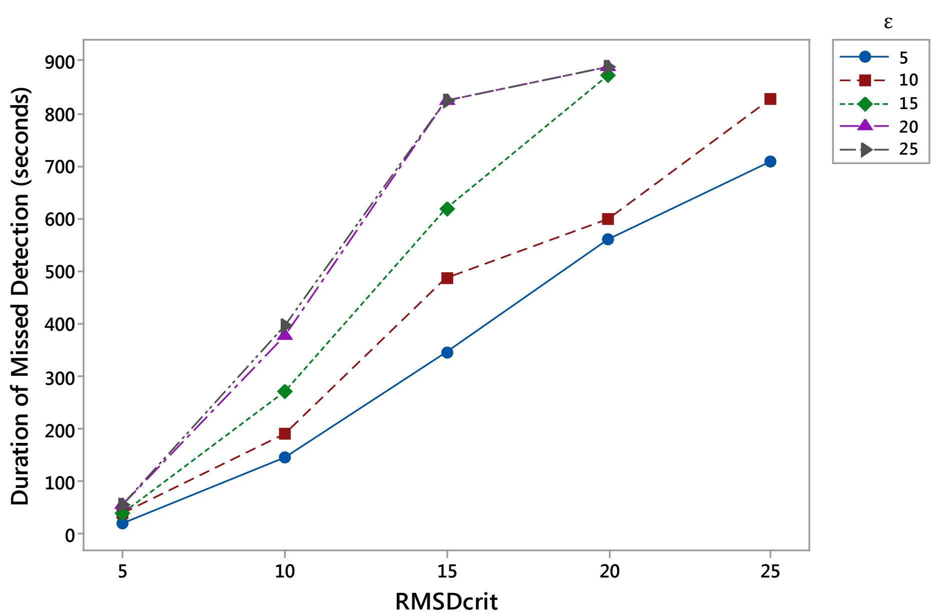


Figure A – 101 Participant 13’s duration of missed detection (seconds) vs. $RMSD_{crit}$= (5, 10, 15, 20, and 25)

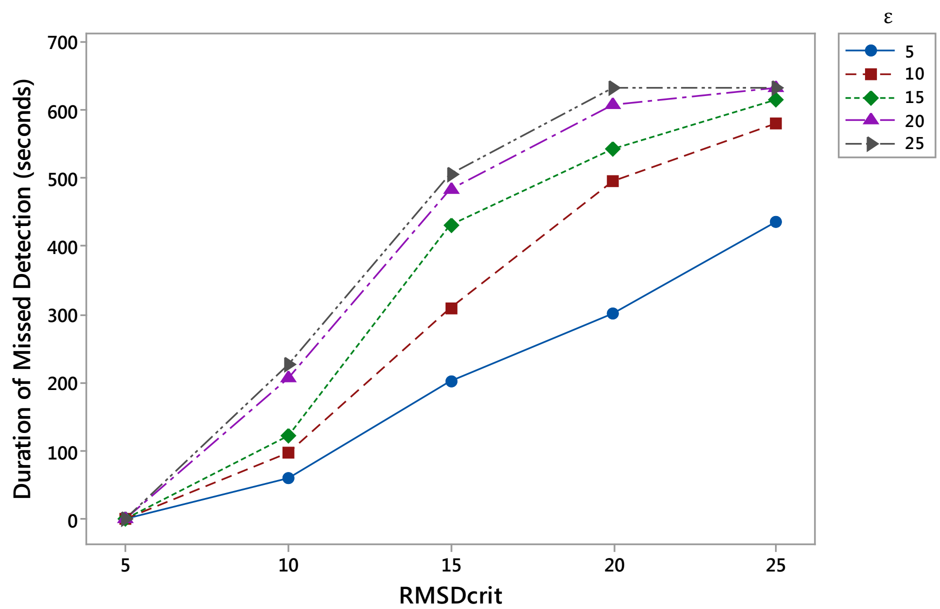


Figure A – 102 Participant 14's duration of missed detection (seconds) vs. $RMSD_{crit}$= (5, 10, 15, 20, and 25)

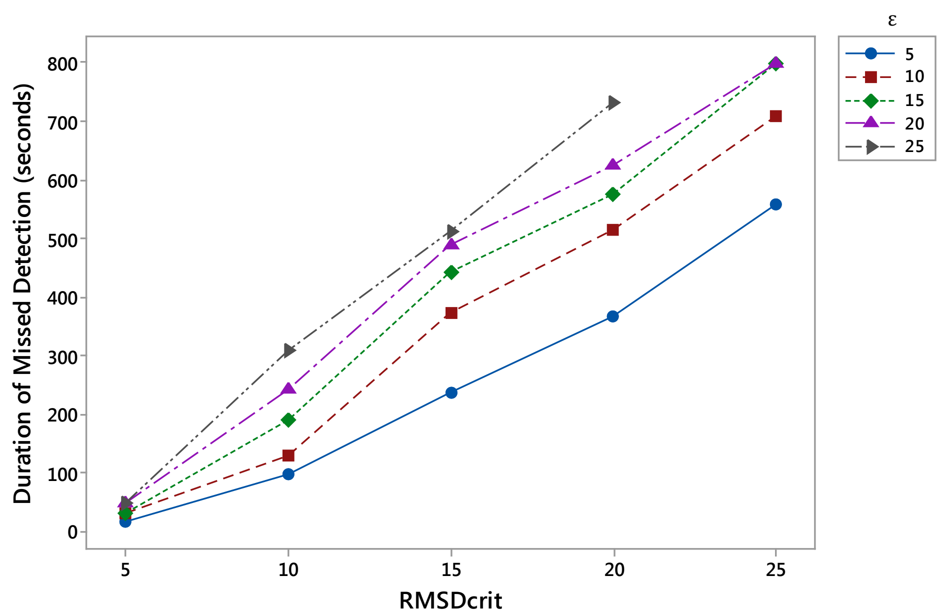


Figure A – 103 Participant 15's duration of missed detection (seconds) vs. $RMSD_{crit}$= (5, 10, 15, 20, and 25)

Table A – 1 Evaluation of parameters ($RMSD_{crit}, \varepsilon$) on participant 1's tracking task performance

$RMSD_{crit}$		ε				
		5	10	15	20	25
5	FALSE (seconds)	551	518.8	518.8	494.4	470.4
	MISS (seconds)	0	0	0	0	0
	DELAY (seconds)	0	0	0	0	0
10	FALSE (seconds)	442	373.5	362.5	329.5	242
	MISS (seconds)	32.8	57.8	57.8	88.7	130.7
	DELAY (seconds)	5	10	15	20	25
15	FALSE (seconds)	316	197.1	185.1	77.9	77.9
	MISS (seconds)	103.7	146.6	232.8	314.9	423
	DELAY (seconds)	0	0	0	98.9	242.9
20	FALSE (seconds)	215.7	128.2	103.2	53	53
	MISS (seconds)	240.3	335.5	435	487	487
	DELAY (seconds)	0	0	0	N/A	N/A
25	FALSE (seconds)	142.5	97.9	26	26	26
	MISS (seconds)	313.9	429.1	452.1	487	487
	DELAY (seconds)	0	0	0	N/A	N/A

Table A – 2 Evaluation of parameters ($RMSD_{crit}, \varepsilon$) on participant 2's tracking task performance

$RMSD_{crit}$		ε				
		5	10	15	20	25
5	FALSE (seconds)	268.1	205.1	192.1	158.1	134.1
	MISS (seconds)	6	11	11	10	35
	DELAY (seconds)	0	17	22	27	32
10	FALSE (seconds)	113.6	55.9	32.9	0	0
	MISS (seconds)	192	260.4	354.3	428.2	471.2
	DELAY (seconds)	31	45.1	50.1	90	95
15	FALSE (seconds)	23	0	0	0	0
	MISS (seconds)	380.1	529.1	644.2	659.2	700.2
	DELAY (seconds)	31	93	513.1	518.1	644.1
20	FALSE (seconds)	0	0	0	0	0
	MISS (seconds)	568.2	689.1	1080	1080	1080
	DELAY (seconds)	81	292.1	N/A	N/A	N/A
25	FALSE (seconds)	0	0	0	0	0
	MISS (seconds)	665	1080	1080	1080	1080
	DELAY (seconds)	N/A	N/A	N/A	N/A	N/A

Table A – 3 Evaluation of parameters ($RMSD_{crit}, \varepsilon$) on participant 4's tracking task performance

$RMSD_{crit}$		ε				
		5	10	15	20	25
5	FALSE (seconds)	323	323	323	323	323
	MISS (seconds)	1.9	1.9	1.9	1.9	1.9
	DELAY (seconds)	0	0	0	0	0
10	FALSE (seconds)	249.5	231.4	218.4	183.3	91.1
	MISS (seconds)	83.7	124.61	193.5	233.7	365
	DELAY (seconds)	3.1	18.9	23.9	28.9	74
15	FALSE (seconds)	139	93.9	18	0	0
	MISS (seconds)	264	371.6	501.7	568.8	671
	DELAY (seconds)	18	23	28	33	N/A
20	FALSE (seconds)	78.8	23	0	0	0
	MISS (seconds)	406.8	530.9	688.9	738	738
	DELAY (seconds)	21	26	31	N/A	N/A
25	FALSE (seconds)	20.9	0	0	0	0
	MISS (seconds)	554	567.2	740	740	740
	DELAY (seconds)	24	29	N/A	N/A	N/A

Table A – 4 Evaluation of parameters ($RMSD_{crit}, \varepsilon$) on participant 5's tracking task performance

$RMSD_{crit}$		ε				
		5	10	15	20	25
5	FALSE (seconds)	306.7	281.6	239.6	204.6	204.6
	MISS (seconds)	7.1	7.1	19.1	34.2	34.2
	DELAY (seconds)	0	3.1	8.1	13.1	18.1
10	FALSE (seconds)	167.2	140.2	117.2	101.1	72.9
	MISS (seconds)	99.1	149.2	175.1	194.1	212.2
	DELAY (seconds)	0	6	9	14	121
15	FALSE (seconds)	107.2	47.1	15	0	0
	MISS (seconds)	175.4	341	341	395.1	437.2
	DELAY (seconds)	0	56.9	158.1	163.1	168.1
20	FALSE (seconds)	73	25.9	0	0	0
	MISS (seconds)	330.1	433.8	560.7	579.7	645
	DELAY (seconds)	2.1	79.1	402	407	456
25	FALSE (seconds)	31.9	0	0	0	0
	MISS (seconds)	470.3	553	579	624.9	644.9
	DELAY (seconds)	1.9	400.1	405.1	451	456

Table A – 5 Evaluation of parameters ($RMSD_{crit}, \varepsilon$) on participant 6's tracking task performance

$RMSD_{crit}$		ε				
		5	10	15	20	25
5	FALSE (seconds)	880.1	880.1	880.1	880.1	880.1
	MISS (seconds)	0	0	0	0	0
	DELAY (seconds)	0	0	0	0	0
10	FALSE (seconds)	817.7	797.7	772.8	711.7	668.1
	MISS (seconds)	10	10	10	16.9	38
	DELAY (seconds)	0	0	2	28.9	62
15	FALSE (seconds)	702.5	593	507	390.1	342
	MISS (seconds)	27.3	32.3	49.2	49.2	69.2
	DELAY (seconds)	1.9	6.9	23.9	28.9	75
20	FALSE (seconds)	543.3	407.5	288.5	190.7	166.7
	MISS (seconds)	46.2	72.1	98.2	98.2	98.2
	DELAY (seconds)	0	0	73	78	83
25	FALSE (seconds)	397.3	221.6	124.7	55	32
	MISS (seconds)	68.4	117.1	158.1	158.1	158.1
	DELAY (seconds)	1.9	1.9	77	82	87

Table A – 6 Evaluation of parameters ($RMSD_{crit}, \varepsilon$) on participant 7's tracking task performance

$RMSD_{crit}$		ε				
		5	10	15	20	25
5	FALSE (seconds)	265	265	265	265	265
	MISS (seconds)	3.9	12.8	12.8	12.8	12.8
	DELAY (seconds)	0	0	0	0	0
10	FALSE (seconds)	198.3	156.1	140.6	123.5	103.4
	MISS (seconds)	69.8	84.7	178.7	231.8	301.7
	DELAY (seconds)	0	0	8	13	18
15	FALSE (seconds)	96.2	62.1	25	25	25
	MISS (seconds)	214.5	281.5	389.8	428.8	519.8
	DELAY (seconds)	0	0	3	155.1	482.1
20	FALSE (seconds)	38	5	0	0	0
	MISS (seconds)	341.1	499.1	582.2	670.1	690.1
	DELAY (seconds)	0	5	265.1	270.1	557.1
25	FALSE (seconds)	7.1	0	0	0	0
	MISS (seconds)	450.3	602.2	742	775.1	N/A
	DELAY (seconds)	4	260.1	680.1	780.1	N/A

Table A – 7 Evaluation of parameters ($RMSD_{crit}, \varepsilon$) on participant 10's tracking task performance

$RMSD_{crit}$		ε				
		5	10	15	20	25
5	FALSE (seconds)	379.9	375	375	375	375
	MISS (seconds)	2.1	2.1	2.1	2.1	2.1
	DELAY (seconds)	0	0	0	0	0
10	FALSE (seconds)	336.4	320.5	320.5	320.5	320.5
	MISS (seconds)	31.3	54.2	54.2	54.2	54.2
	DELAY (seconds)	0	0	0	0	0
15	FALSE (seconds)	261.8	245.9	245.9	211.9	211.9
	MISS (seconds)	86.1	123	185.1	204.1	204.1
	DELAY (seconds)	0	0	0	0	0
20	FALSE (seconds)	218.6	192.8	192.8	159.9	115.9
	MISS (seconds)	137.8	191.8	307.9	342	414
	DELAY (seconds)	0	0	0	0	0
25	FALSE (seconds)	159.9	116.9	93.9	77	77
	MISS (seconds)	191.8	269	354.9	422	513.1
	DELAY (seconds)	0	0	0	0	0

Table A – 8 Evaluation of parameters ($RMSD_{crit}, \varepsilon$) on participant 11's tracking task performance

$RMSD_{crit}$		ε				
		5	10	15	20	25
5	FALSE (seconds)	655.7	655.7	613.7	596.8	596.8
	MISS (seconds)	0	0	0	0	0
	DELAY (seconds)	0	0	0	0	0
10	FALSE (seconds)	471.9	367.7	252.6	170.6	125.5
	MISS (seconds)	33.8	55.8	79.9	95.9	95.9
	DELAY (seconds)	0	0	0	0	0
15	FALSE (seconds)	248.8	144.6	72.4	47.5	26
	MISS (seconds)	117.9	176.9	213	310.1	354
	DELAY (seconds)	0	1.1	6.1	133	138
20	FALSE (seconds)	120.8	48.9	34	0	0
	MISS (seconds)	240.1	312	347.1	383.1	383.1
	DELAY (seconds)	5	97	290.1	N/A	N/A
25	FALSE (seconds)	63.9	26.9	0	0	
	MISS (seconds)	302.1	371	396.1	396.1	396.1
	DELAY (seconds)	6	124	N/A	N/A	N/A

Table A –9 Evaluation of parameters ($RMSD_{crit}, \varepsilon$) on participant 12's tracking task performance

$RMSD_{crit}$		ε				
		5	10	15	20	25
5	FALSE (seconds)	918.6	913.6	900.6	900.6	900.6
	MISS (seconds)	0	0	0	0	0
	DELAY (seconds)	0	0	0	0	0
10	FALSE (seconds)	710.9	591.2	468.5	355.4	284.2
	MISS (seconds)	1.1	1.1	1.1	1.1	1.1
	DELAY (seconds)	0	0	3	8	13
15	FALSE (seconds)	424.4	230.4	158.1	86.1	64.1
	MISS (seconds)	7	7	7	26	26
	DELAY (seconds)	0	0	3	8	13
20	FALSE (seconds)	246.5	59.2	26	7	0
	MISS (seconds)	11	18	30	49	61.1
	DELAY (seconds)	0	3	8	13	18
25	FALSE (seconds)	111.1	10.1	0	0	0
	MISS (seconds)	26	41	71.1	90.1	111.1
	DELAY (seconds)	0	4.9	9.9	14.9	19.9

Table A – 10 Evaluation of parameters ($RMSD_{crit}, \varepsilon$) on participant 13's tracking task performance

$RMSD_{crit}$		ε				
		5	10	15	20	25
5	FALSE (seconds)	174	166.1	153.1	153.1	153.1
	MISS (seconds)	19.2	38.1	38.1	53.5	53.5
	DELAY (seconds)	0	0	0	0	0
10	FALSE (seconds)	129.6	92.8	67.9	67.9	67.9
	MISS (seconds)	142.9	190.1	270.8	375.9	397
	DELAY (seconds)	0	0	3	8	13
15	FALSE (seconds)	75.1	10.9	0	0	0
	MISS (seconds)	344.1	485.1	618	825.7	825.7
	DELAY (seconds)	5	15	20	*	*
20	FALSE (seconds)	21.1	0	0	0	0
	MISS (seconds)	560.5	597.6	871.5	888.5	888.5
	DELAY (seconds)	14.1	251.6	794.6	N/A	N/A
25	FALSE (seconds)	6	0	0	0	0
	MISS (seconds)	709.7	827.5	N/A	N/A	N/A
	DELAY (seconds)	201.6	251.6	0	0	0

Table A – 11 Evaluation of parameters ($RMSD_{crit}, \varepsilon$) on participant 14's tracking task performance

$RMSD_{crit}$		ε				
		5	10	15	20	25
5	FALSE (seconds)	421.8	412.7	390.3	390.3	390.3
	MISS (seconds)	0	0	0	0	0
	DELAY (seconds)	0	0	0	0	0
10	FALSE (seconds)	321.7	250.6	202	186.9	186.9
	MISS (seconds)	59.7	95.8	121.9	206.1	227.1
	DELAY (seconds)	0.1	9.9	14.9	19.9	24.9
15	FALSE (seconds)	217.9	114.6	65	65	25
	MISS (seconds)	202.1	307.9	431.9	482.8	506.8
	DELAY (seconds)	0	39.9	201.9	314.9	344.9
20	FALSE (seconds)	115.9	46	0	0	0
	MISS (seconds)	302.3	496.1	542.1	609	633
	DELAY (seconds)	0	196.9	201.9	497.9	N/A
25	FALSE (seconds)	29	0	0	0	0
	MISS (seconds)	435.8	580.9	615	633	633
	DELAY (seconds)	0	196.6	196.6	N/A	N/A

Table A – 12 Evaluation of parameters ($RMSD_{crit}, \varepsilon$) on participant 15's tracking task performance

$RMSD_{crit}$		ε				
		5	10	15	20	25
5	FALSE (seconds)	241.6	241.6	216.7	216.7	151.1
	MISS (seconds)	15.9	28.9	28.9	47.9	47.9
	DELAY (seconds)	0	0	0	0	0
10	FALSE (seconds)	156.5	129.2	103.1	84.1	84.1
	MISS (seconds)	95.8	128	188.1	241.2	308.2
	DELAY (seconds)	0	0	0	0	1
15	FALSE (seconds)	110.2	78.1	55.1	37.2	16.1
	MISS (seconds)	235.3	372.6	441.8	489.5	510.6
	DELAY (seconds)	0	0	0	3.9	8.9
20	FALSE (seconds)	76.2	35.1	35.1	16.1	16.1
	MISS (seconds)	365.2	513.5	574.5	624.4	732.6
	DELAY (seconds)	0	0	0	3.9	8.9
25	FALSE (seconds)	49.1	31	19	0	0
	MISS (seconds)	558.3	708.3	798.3	798.3	N/A
	DELAY (seconds)	0	0	446.4	451.4	N/A

Table A – 13 Evaluation of parameters ($RMSD_{crit}, \varepsilon$) on participant 1's tracking task performance

$RMSD_{crit}$ (pixel units)		ε				
		5	10	15	20	25
5	1-Specificity	0.95	0.89	0.89	0.85	0.81
	Sensitivity	1.00	1.00	1.00	1.00	1.00
10	1-Specificity	0.76	0.64	0.62	0.57	0.42
	Sensitivity	0.95	0.92	0.92	0.88	0.84
15	1-Specificity	0.54	0.34	0.32	0.13	0.13
	Sensitivity	0.86	0.83	0.74	0.69	0.58
20	1-Specificity	0.37	0.22	0.18	0.09	0.09
	Sensitivity	0.72	0.65	0.55	0.53	0.53
25	1-Specificity	0.25	0.17	0.04	0.04	0.04
	Sensitivity	0.67	0.56	0.57	0.54	0.54

Table A – 14 Evaluation of parameters ($RMSD_{crit}, \varepsilon$) on participant 2's tracking task performance

$RMSD_{crit}$ (pixel units)		ε				
		5	10	15	20	25
5	1-Specificity	0.81	0.62	0.58	0.48	0.41
	Sensitivity	0.99	0.99	0.99	0.99	0.96
10	1-Specificity	0.34	0.17	0.10	0.00	0.00
	Sensitivity	0.80	0.75	0.66	0.60	0.56
15	1-Specificity	0.07	0.00	0.00	0.00	0.00
	Sensitivity	0.64	0.51	0.40	0.39	0.35
20	1-Specificity	0.00	0.00	0.00	0.00	0.00
	Sensitivity	0.47	0.36	0.00	0.00	0.00
25	1-Specificity	0.00	0.00	0.00	0.00	0.00
	Sensitivity	0.38	0.00	0.00	0.00	0.00

Table A – 15 Evaluation of parameters ($RMSD_{crit}, \varepsilon$) on participant 4's tracking task performance

$RMSD_{crit}$ (pixel units)		ε				
		5	10	15	20	25
5	1-Specificity	0.98	0.98	0.98	0.98	0.98
	Sensitivity	1.00	1.00	1.00	1.00	1.00
10	1-Specificity	0.75	0.70	0.66	0.55	0.28
	Sensitivity	0.90	0.85	0.78	0.74	0.63
15	1-Specificity	0.42	0.28	0.05	0.00	0.00
	Sensitivity	0.72	0.62	0.53	0.47	0.38
20	1-Specificity	0.24	0.07	0.00	0.00	0.00
	Sensitivity	0.59	0.50	0.36	0.32	0.32
25	1-Specificity	0.06	0.00	0.00	0.00	0.00
	Sensitivity	0.48	0.47	0.31	0.31	0.31

Table A – 16 Evaluation of parameters ($RMSD_{crit}, \varepsilon$) on participant 5's tracking task performance

$RMSD_{crit}$ (pixel units)		ε				
		5	10	15	20	25
5	1-Specificity	0.77	0.71	0.60	0.52	0.52
	Sensitivity	0.99	0.99	0.98	0.96	0.96
10	1-Specificity	0.42	0.35	0.30	0.25	0.18
	Sensitivity	0.89	0.84	0.82	0.80	0.79
15	1-Specificity	0.27	0.12	0.04	0.00	0.00
	Sensitivity	0.82	0.67	0.68	0.63	0.60
20	1-Specificity	0.18	0.07	0.00	0.00	0.00
	Sensitivity	0.67	0.59	0.48	0.46	0.40
25	1-Specificity	0.08	0.00	0.00	0.00	0.00
	Sensitivity	0.55	0.49	0.46	0.42	0.40

Table A – 17 Evaluation of parameters ($RMSD_{crit}, \varepsilon$) on participant 6's tracking task performance

$RMSD_{crit}$ (pixel units)		ε				
		5	10	15	20	25
5	1-Specificity	0.99	0.99	0.99	0.99	0.99
	Sensitivity	1.00	1.00	1.00	1.00	1.00
10	1-Specificity	0.92	0.90	0.87	0.80	0.75
	Sensitivity	0.96	0.96	0.96	0.95	0.91
15	1-Specificity	0.79	0.67	0.57	0.44	0.39
	Sensitivity	0.93	0.93	0.91	0.93	0.91
20	1-Specificity	0.61	0.46	0.33	0.22	0.19
	Sensitivity	0.91	0.89	0.88	0.89	0.89
25	1-Specificity	0.45	0.25	0.14	0.06	0.04
	Sensitivity	0.90	0.86	0.83	0.85	0.85

Table A – 18 Evaluation of parameters ($RMSD_{crit}, \varepsilon$) on participant 7's tracking task performance

$RMSD_{crit}$ (pixel units)		ε				
		5	10	15	20	25
5	1-Specificity	0.96	0.96	0.96	0.96	0.96
	Sensitivity	1.00	0.98	0.98	0.98	0.98
10	1-Specificity	0.72	0.57	0.51	0.45	0.37
	Sensitivity	0.92	0.91	0.81	0.76	0.69
15	1-Specificity	0.35	0.23	0.09	0.09	0.09
	Sensitivity	0.78	0.72	0.63	0.59	0.51
20	1-Specificity	0.14	0.02	0.00	0.00	0.00
	Sensitivity	0.67	0.54	0.46	0.38	0.36
25	1-Specificity	0.03	0.00	0.00	0.00	0.00
	Sensitivity	0.58	0.44	0.31	0.28	0.00

Table A – 19 Evaluation of parameters ($RMSD_{crit}, \varepsilon$) on participant 10's tracking task performance

$RMSD_{crit}$ (pixel units)		ε				
		5	10	15	20	25
5	1-Specificity	0.97	0.96	0.96	0.96	0.96
	Sensitivity	1.00	1.00	1.00	1.00	1.00
10	1-Specificity	0.86	0.82	0.82	0.82	0.82
	Sensitivity	0.96	0.93	0.93	0.93	0.93
15	1-Specificity	0.67	0.63	0.63	0.54	0.54
	Sensitivity	0.89	0.85	0.78	0.76	0.76
20	1-Specificity	0.56	0.49	0.49	0.41	0.30
	Sensitivity	0.84	0.78	0.65	0.63	0.57
25	1-Specificity	0.41	0.30	0.24	0.20	0.20
	Sensitivity	0.79	0.72	0.64	0.58	0.49

Table A – 20 Evaluation of parameters ($RMSD_{crit}, \varepsilon$) on participant 11's tracking task performance

$RMSD_{crit}$ (pixel units)		ε				
		5	10	15	20	25
5	1-Specificity	0.96	0.96	0.90	0.88	0.88
	Sensitivity	1.00	1.00	1.00	1.00	1.00
10	1-Specificity	0.69	0.54	0.37	0.25	0.18
	Sensitivity	0.94	0.92	0.90	0.89	0.90
15	1-Specificity	0.37	0.21	0.11	0.07	0.04
	Sensitivity	0.86	0.81	0.79	0.70	0.66
20	1-Specificity	0.18	0.07	0.05	0.00	0.00
	Sensitivity	0.75	0.70	0.67	0.65	0.65
25	1-Specificity	0.09	0.04	0.00	0.00	0.00
	Sensitivity	0.70	0.65	0.63	0.63	0.63

Table A – 21 Evaluation of parameters ($RMSD_{crit}, \varepsilon$) on participant 12's tracking task performance

$RMSD_{crit}$ (pixel units)		ε				
		5	10	15	20	25
5	1-Specificity	0.95	0.95	0.93	0.93	0.93
	Sensitivity	1.00	1.00	1.00	1.00	1.00
10	1-Specificity	0.74	0.61	0.49	0.37	0.29
	Sensitivity	1.00	1.00	1.00	1.00	1.00
15	1-Specificity	0.44	0.24	0.16	0.09	0.07
	Sensitivity	0.99	0.99	0.99	0.97	0.97
20	1-Specificity	0.26	0.06	0.03	0.01	0.00
	Sensitivity	0.99	0.98	0.97	0.95	0.94
25	1-Specificity	0.12	0.01	0.00	0.00	0.00
	Sensitivity	0.97	0.96	0.93	0.92	0.90

Table A – 22 Evaluation of parameters ($RMSD_{crit}, \varepsilon$) on participant 13's tracking task performance

$RMSD_{crit}$ (pixel units)		ε				
		5	10	15	20	25
5	1-Specificity	0.95	0.91	0.84	0.84	0.84
	Sensitivity	0.98	0.96	0.96	0.94	0.94
10	1-Specificity	0.71	0.51	0.37	0.37	0.37
	Sensitivity	0.85	0.81	0.73	0.63	0.61
15	1-Specificity	0.41	0.06	0.00	0.00	0.00
	Sensitivity	0.66	0.55	0.43	0.24	0.24
20	1-Specificity	0.12	0.00	0.00	0.00	0.00
	Sensitivity	0.47	0.45	0.19	0.18	0.18
25	1-Specificity	0.03	0.00	0.00	0.00	0.00
	Sensitivity	0.34	0.23	0.00	0.00	0.00

Table A – 23 Evaluation of parameters ($RMSD_{crit}, \varepsilon$) on participant 14's tracking task performance

$RMSD_{crit}$ (pixel units)		ε				
		5	10	15	20	25
5	1-Specificity	0.97	0.95	0.90	0.90	0.90
	Sensitivity	1.00	1.00	1.00	1.00	1.00
10	1-Specificity	0.74	0.57	0.46	0.43	0.43
	Sensitivity	0.92	0.88	0.86	0.77	0.75
15	1-Specificity	0.50	0.26	0.15	0.15	0.06
	Sensitivity	0.77	0.68	0.57	0.52	0.52
20	1-Specificity	0.27	0.11	0.00	0.00	0.00
	Sensitivity	0.69	0.52	0.50	0.44	0.41
25	1-Specificity	0.07	0.00	0.00	0.00	0.00
	Sensitivity	0.59	0.46	0.43	0.41	0.41

Table A – 24 Evaluation of parameters ($RMSD_{crit}, \varepsilon$) on participant 15's tracking task performance

$RMSD_{crit}$ (pixel units)		ε				
		5	10	15	20	25
5	1-Specificity	0.55	0.55	0.50	0.50	0.35
	Sensitivity	0.98	0.97	0.97	0.94	0.95
10	1-Specificity	0.36	0.30	0.24	0.19	0.19
	Sensitivity	0.90	0.87	0.81	0.76	0.69
15	1-Specificity	0.25	0.18	0.13	0.09	0.04
	Sensitivity	0.76	0.63	0.57	0.53	0.52
20	1-Specificity	0.17	0.08	0.08	0.04	0.04
	Sensitivity	0.64	0.51	0.45	0.41	0.31
25	1-Specificity	0.11	0.07	0.04	0.00	0.00
	Sensitivity	0.46	0.32	0.25	0.26	N/A

Appendix B STUDY TWO (DATA)

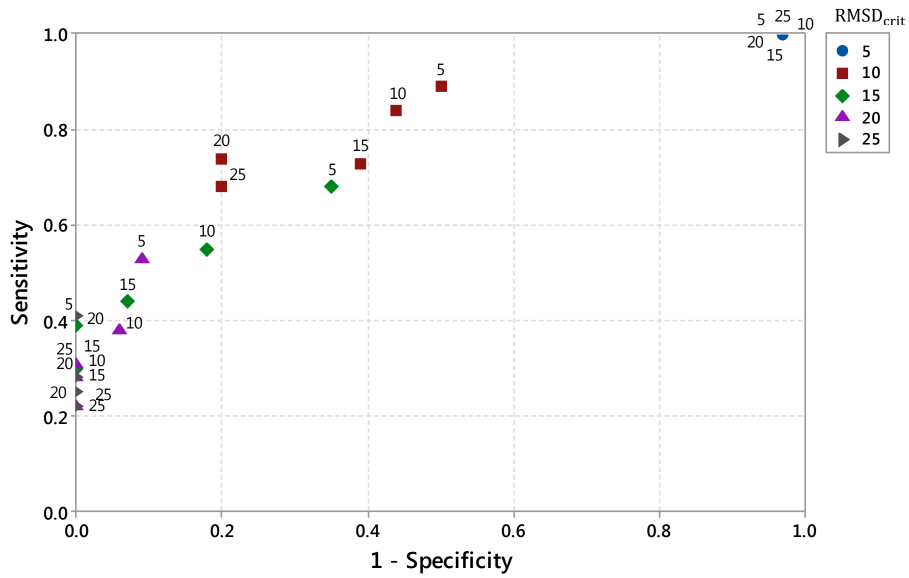


Figure B – 1 ROC curve: evaluation of parameters (RMSDcrit, ε) on participant 1’s tracking task performance

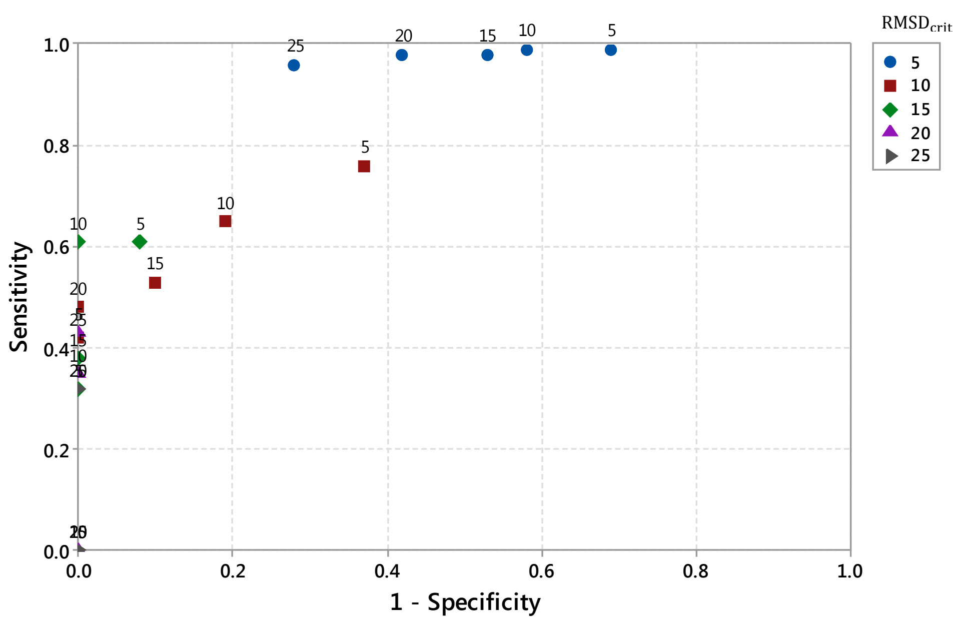


Figure B – 2 ROC curve: evaluation of parameters (RMSDcrit, ε) on participant 2’s tracking task performance

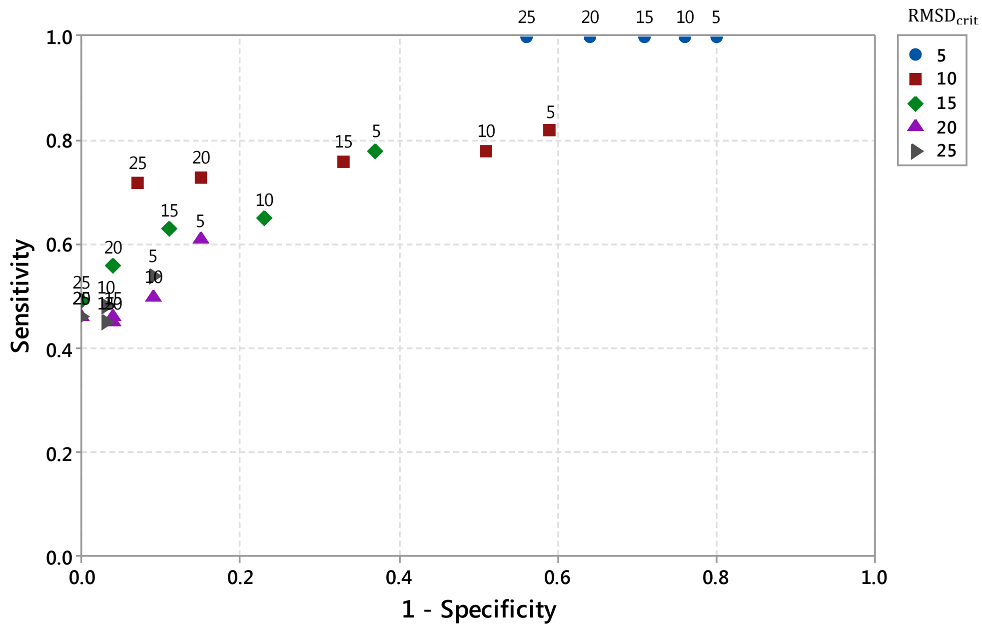


Figure B – 3 ROC curve: evaluation of parameters (RMSDcrit, ε) on participant 4's tracking task performance

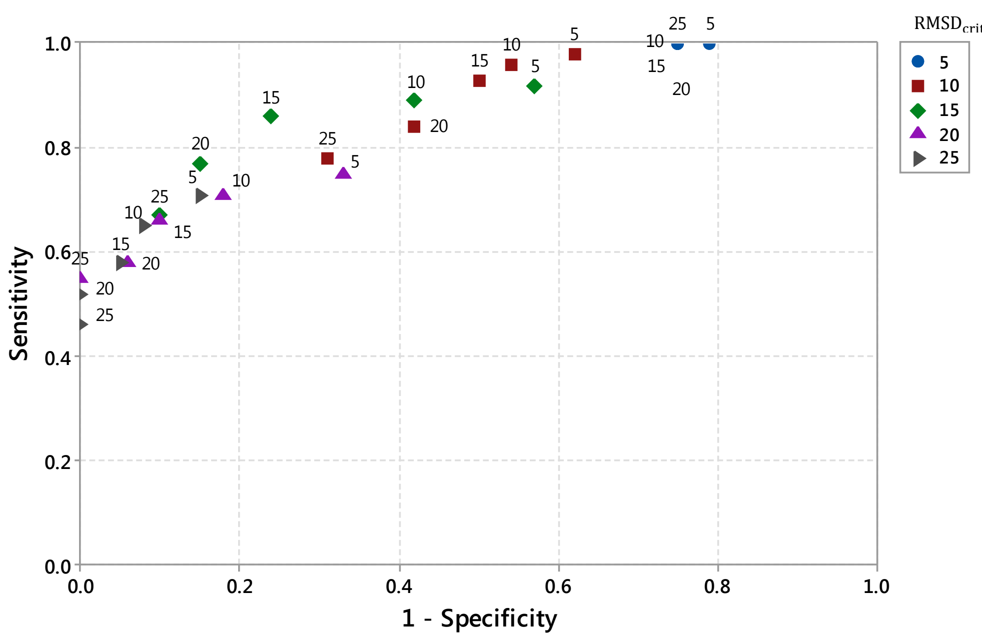


Figure B – 4 ROC curve: evaluation of parameters (RMSDcrit, ε) on participant 5's tracking task performance

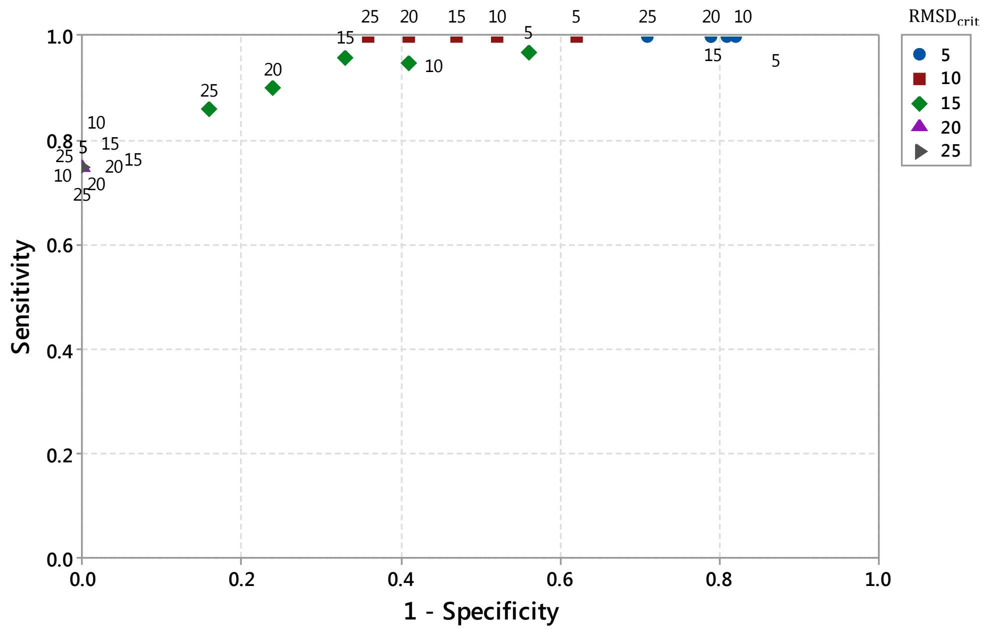


Figure B – 5 ROC curve: evaluation of parameters (RMSDcrit, ε) on participant 6’s tracking task performance

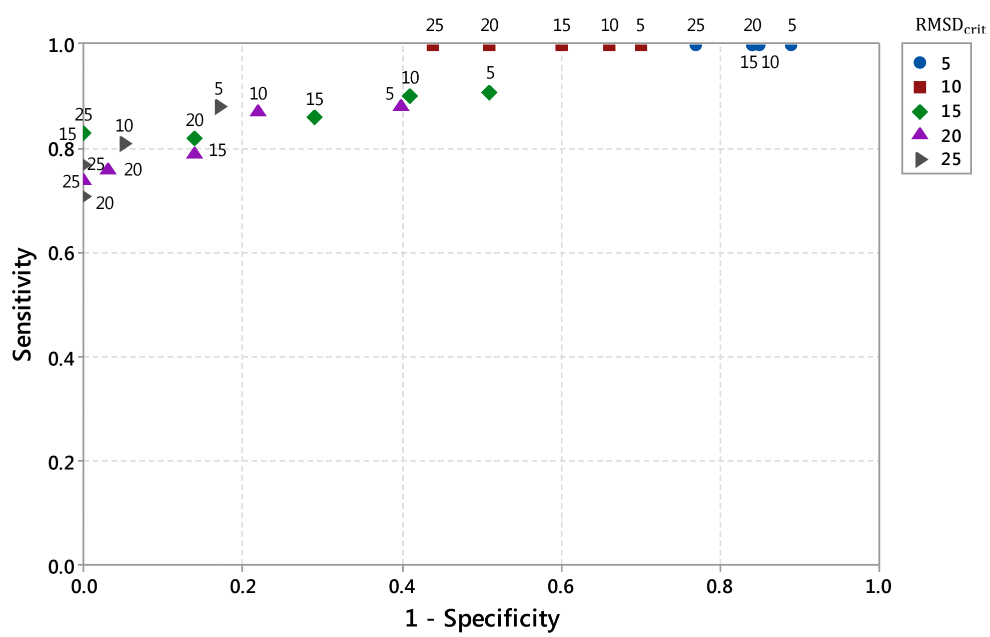


Figure B – 6 ROC curve: evaluation of parameters (RMSDcrit, ε) on participant 7’s tracking task performance

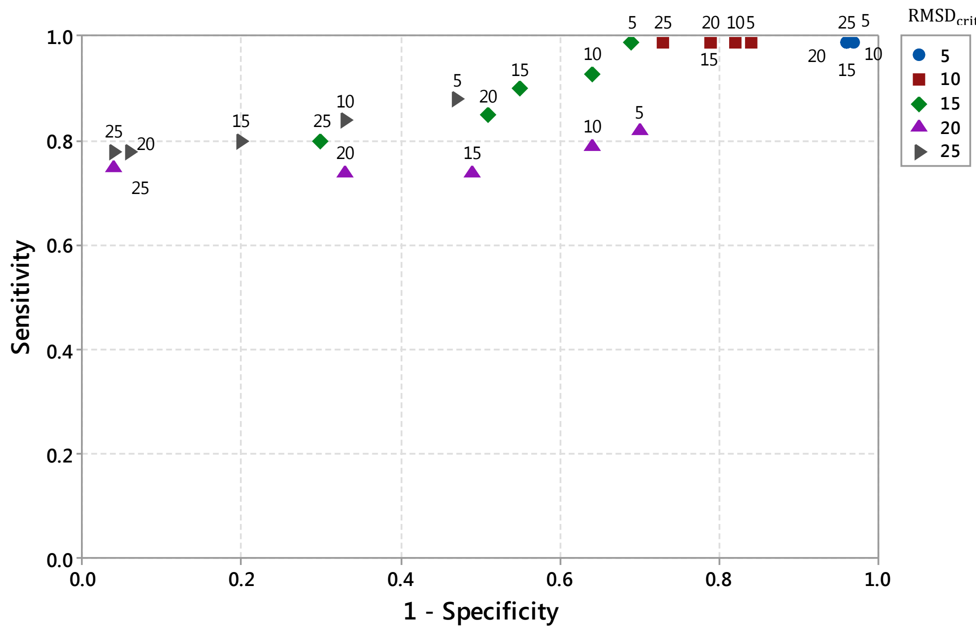


Figure B – 7 ROC curve: evaluation of parameters (RMSDcrit, ε) on participant 10's tracking task performance

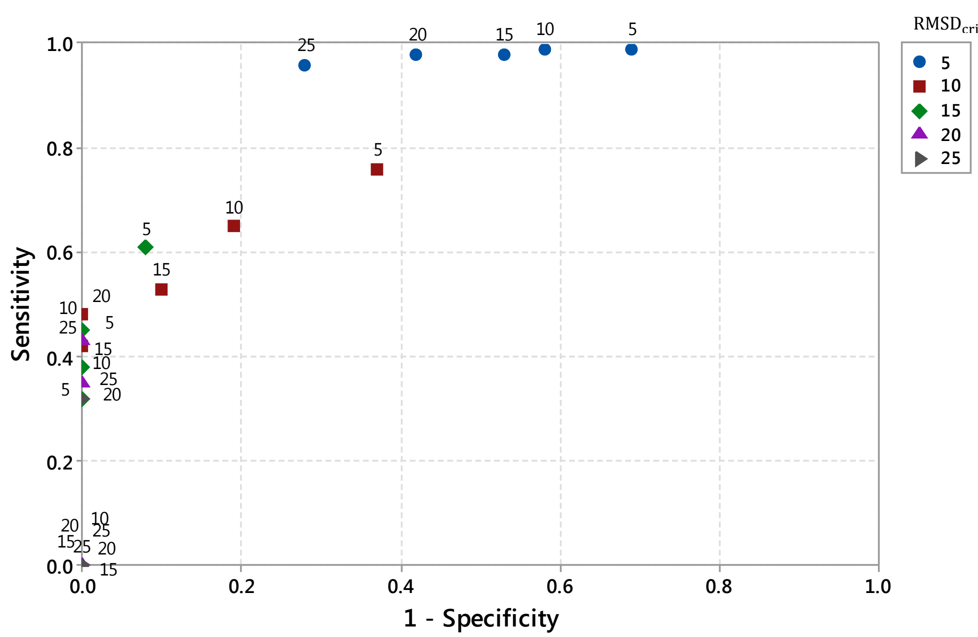


Figure B – 8 ROC curve: evaluation of parameters (RMSDcrit, ε) on participant 11's tracking task performance

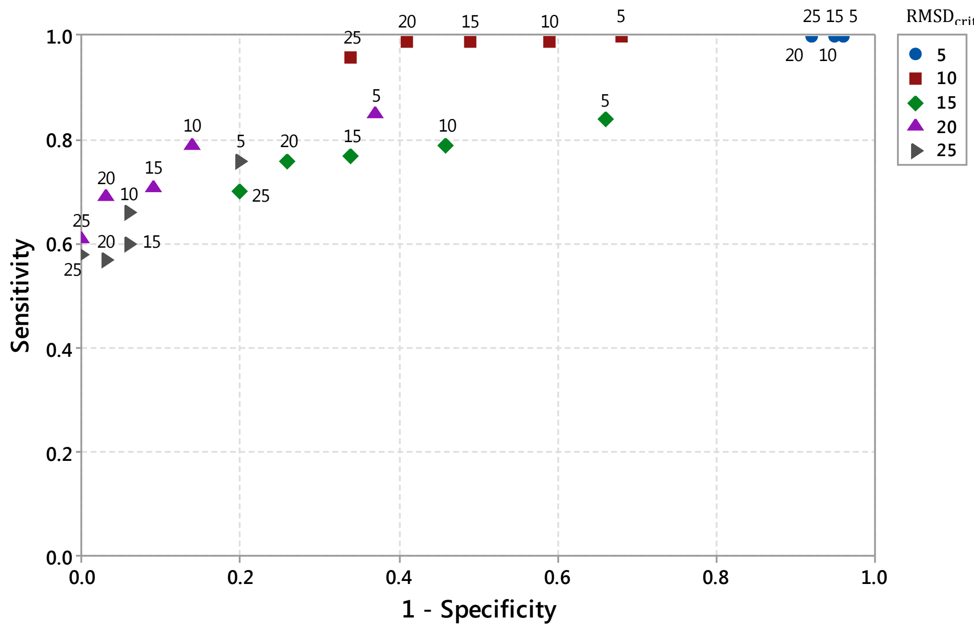


Figure B – 9 ROC curve: evaluation of parameters (RMSDcrit, ε) on participant 12’s tracking task performance

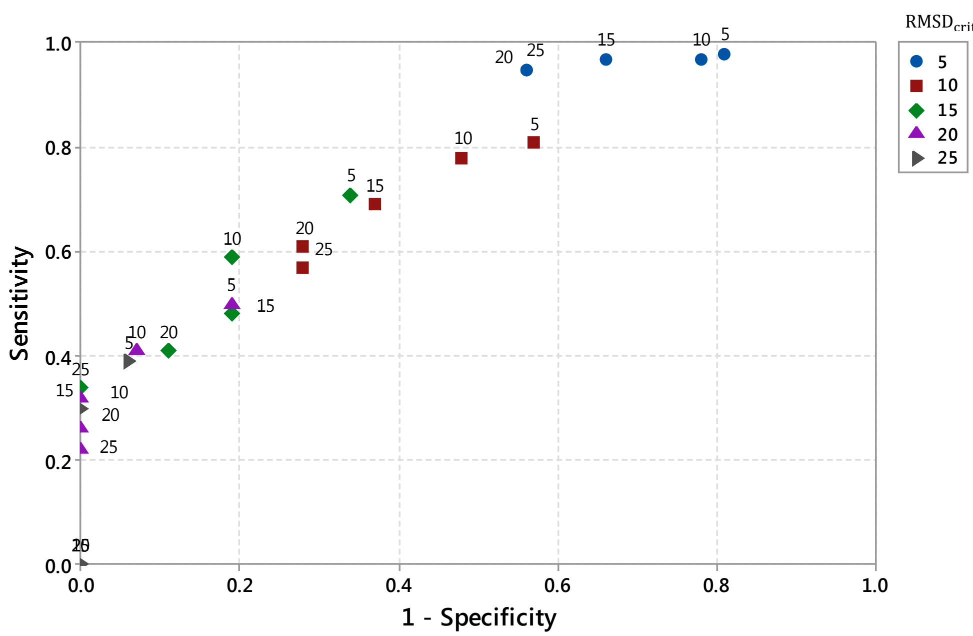


Figure B – 10 ROC curve: evaluation of parameters (RMSDcrit, ε) on participant 13’s tracking task performance

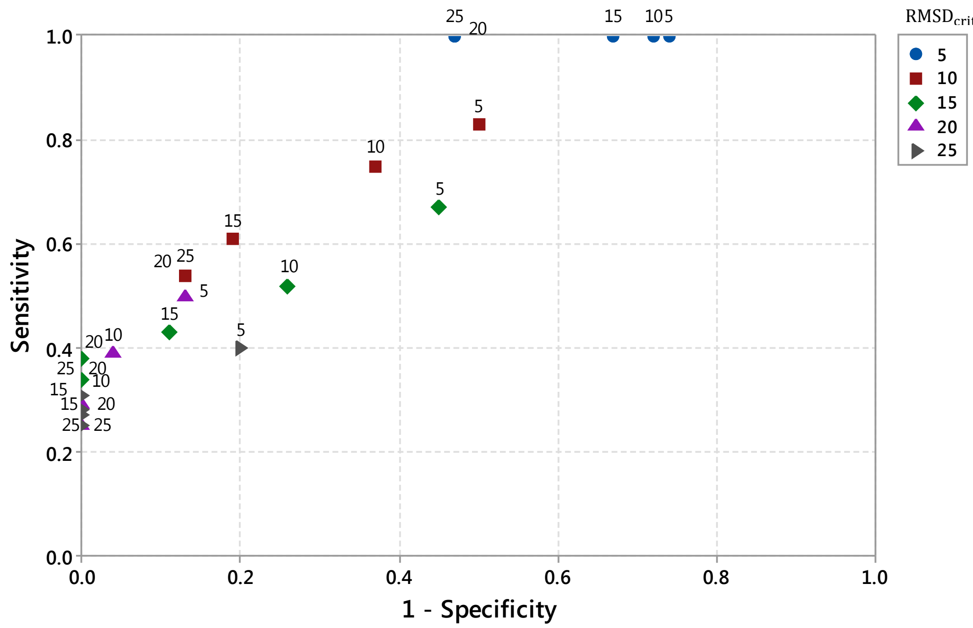


Figure B – 11 ROC curve: evaluation of parameters (RMSDcrit, ε) on participant 14's tracking task performance

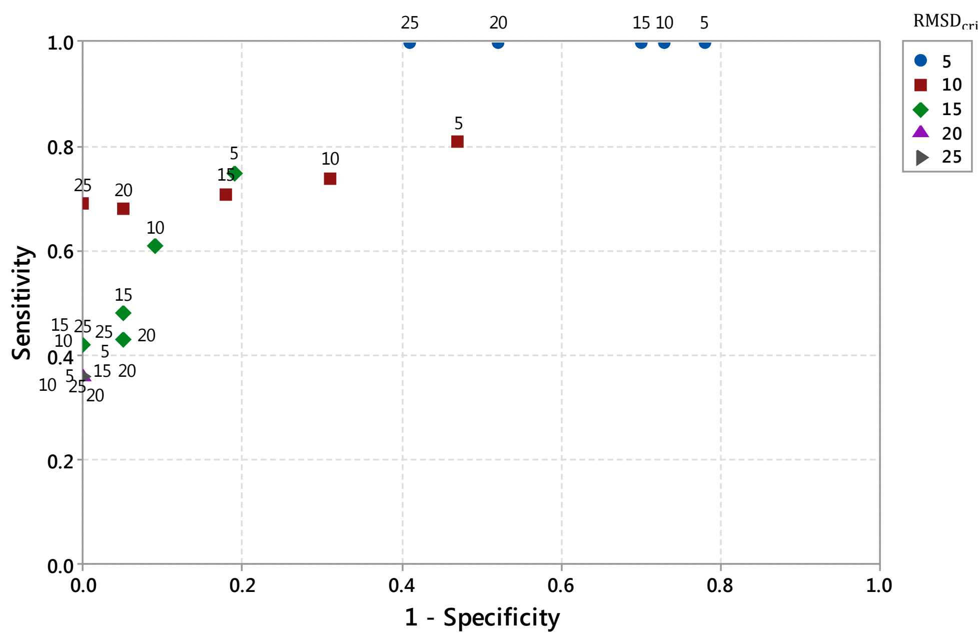


Figure B – 12 ROC curve: evaluation of parameters (RMSDcrit, ε) on participant 15's tracking task performance

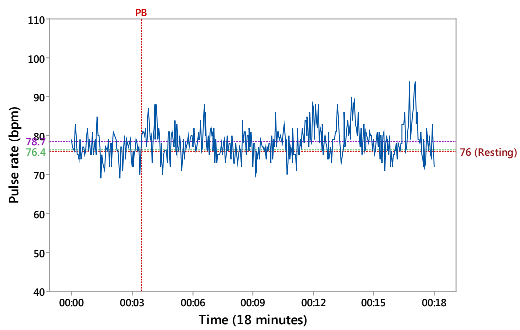


Figure B – 13 Pulse rate (bpm); green = pre-PB, purple = post-PB (participant 1)

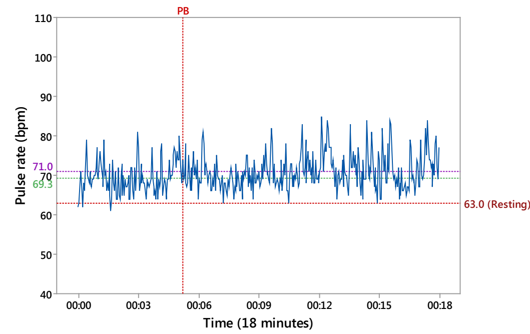


Figure B – 14 Pulse rate (bpm); green = pre-PB, Purple = post-PB (participant 2)

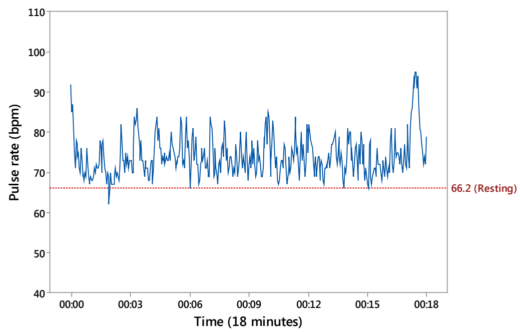


Figure B – 15 Pulse rate (bpm); green = pre-PB, purple = post-PB (participant 3)

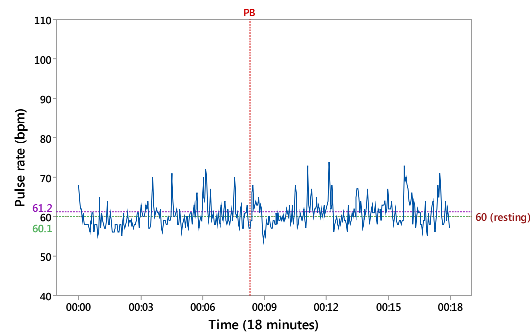


Figure B – 16 Pulse rate (bpm); green = pre-PB, purple = post-PB (participant 4)

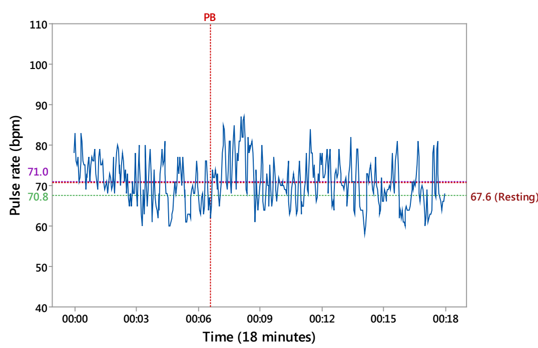


Figure B – 17 Pulse rate (bpm); green = pre-PB, purple = post-PB (participant 5)

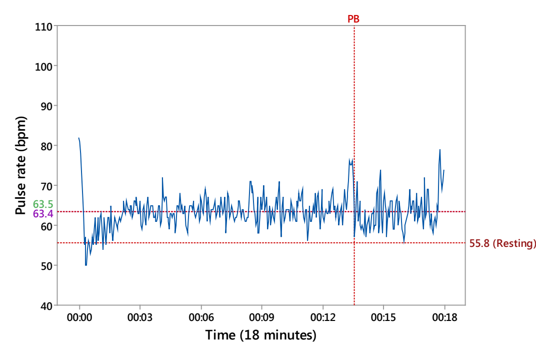


Figure B – 18 Pulse rate (bpm); Green = pre-PB, purple = post-PB (participant 6)

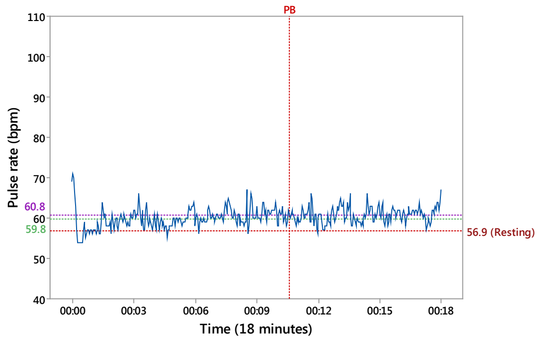


Figure B – 19 Pulse rate (bpm); green = pre-PB, purple = post-PB (Participant 7)

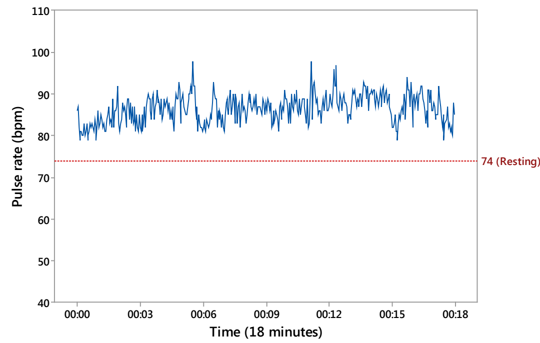


Figure B – 20 Pulse rate (bpm); green = pre-PB, purple = post-PB (participant 8)

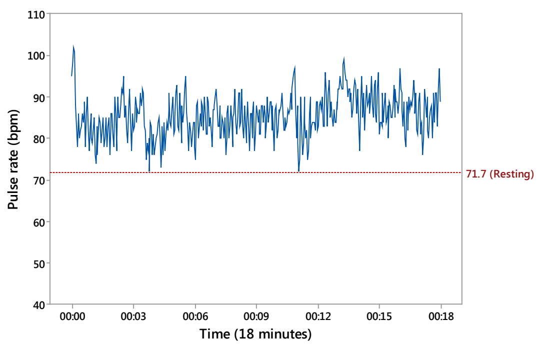


Figure B – 21 Pulse rate (bpm); green = pre-PB, purple = post-PB (participant 9)

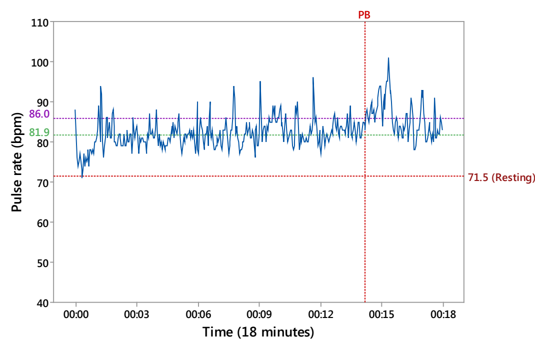


Figure B – 22 Pulse rate (bpm); green = pre-PB, purple = post-PB (participant 10)

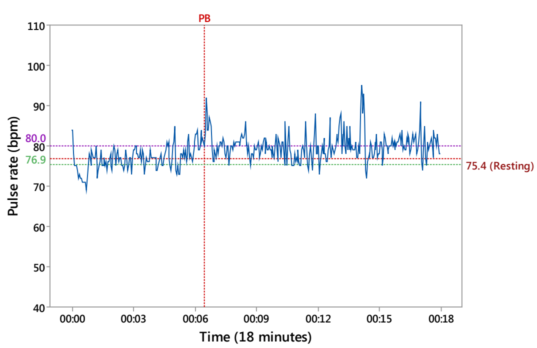


Figure B – 23 Pulse rate (bpm); green = pre-PB, purple = post-PB (participant 11)

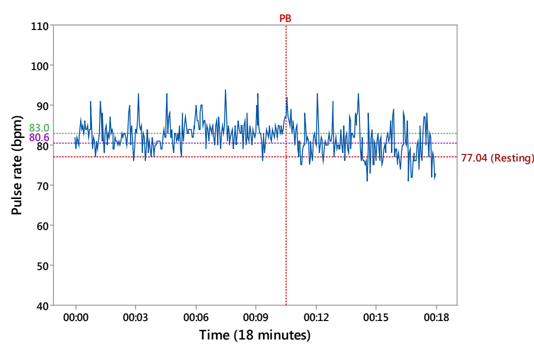


Figure B – 24 Pulse rate (bpm); green = pre-PB, purple = post-PB (participant 12)

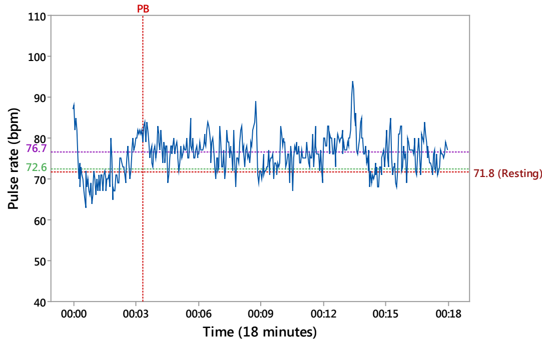


Figure B – 25 Pulse rate (bpm); green = pre-PB, purple = post-PB (participant 13)

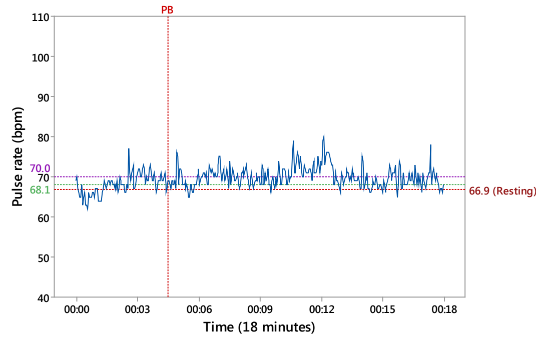


Figure B – 26 Pulse rate (bpm); green = pre-PB, purple = post-PB (participant 14)

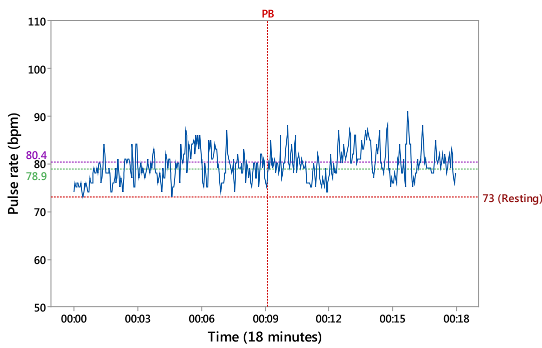


Figure B – 27 Pulse rate (bpm); green = pre-PB, purple = post-PB (participant 15)

Appendix C STUDY TWO VS. STUDY THREE (DATA)

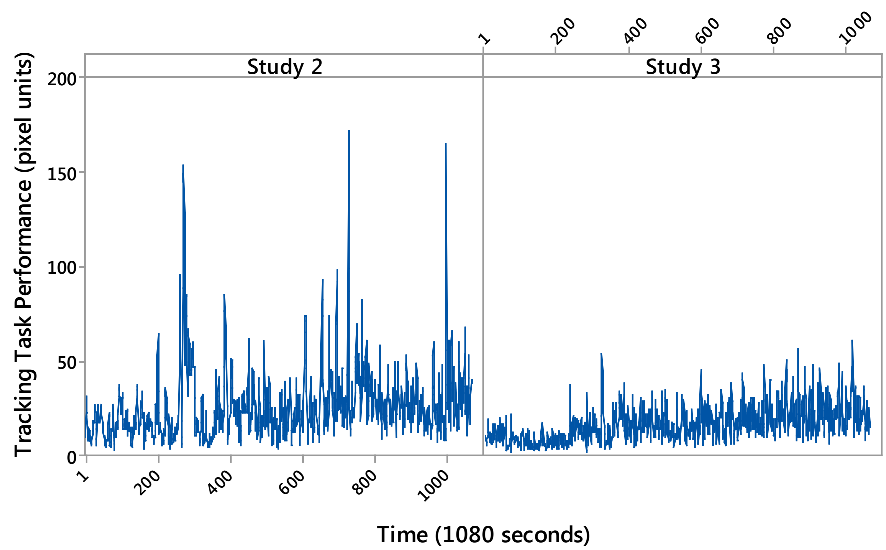


Figure C – 1 Tracking task performance: study 2 vs. study 3 (participant 1)

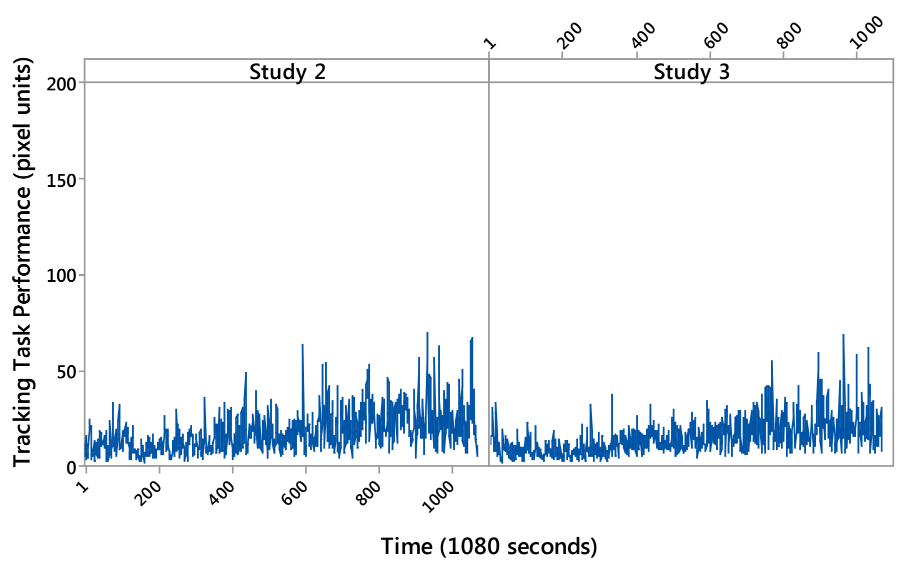


Figure C – 2 Tracking task performance: study 2 vs. study 3 (participant 2)

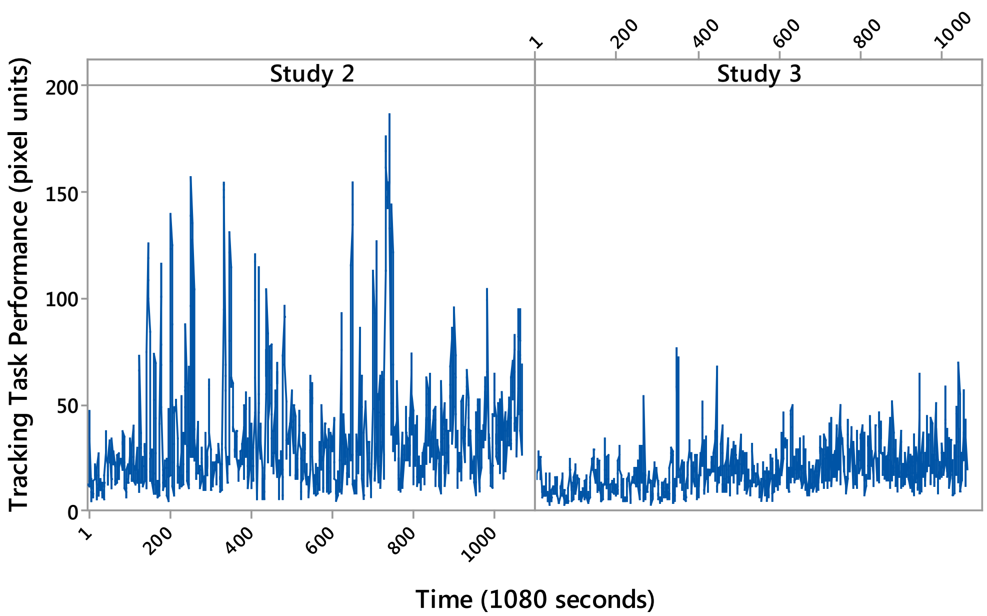


Figure C – 3 Tracking task performance: study 2 vs. study 3 (participant 3)

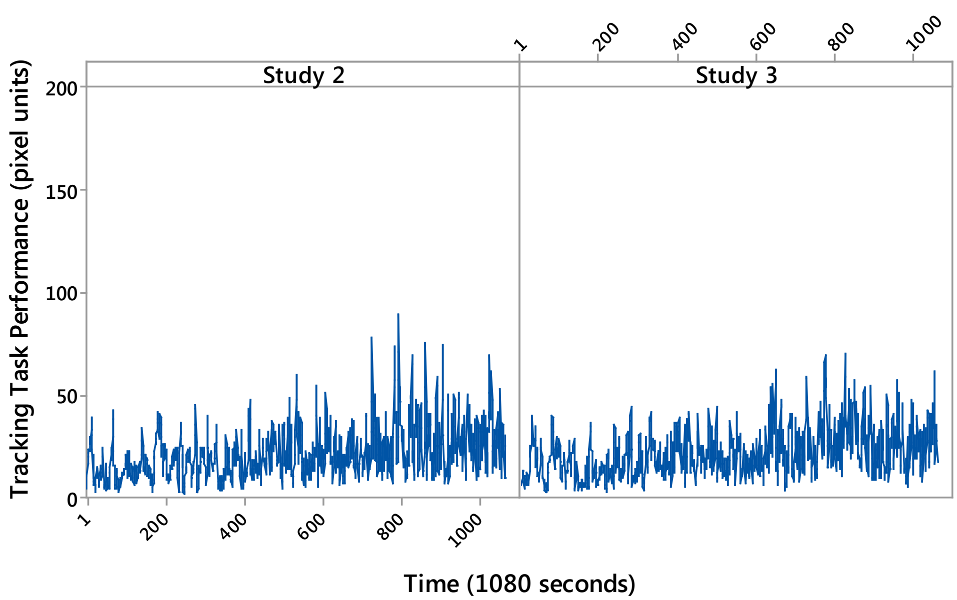


Figure C – 4 Tracking task performance: study 2 vs. study 3 (participant 4)

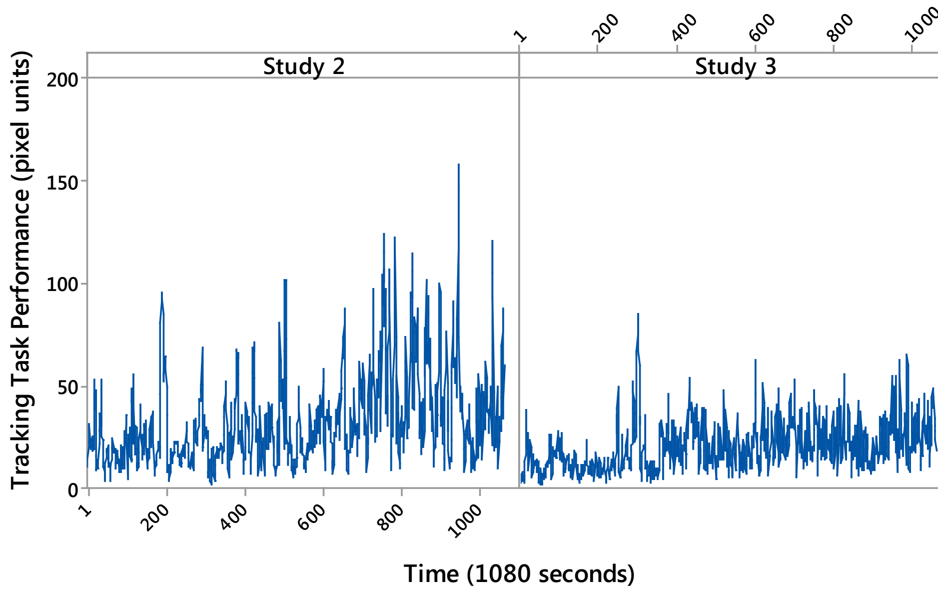


Figure C – 5 Tracking task performance: study 2 vs. study 3 (participant 5)

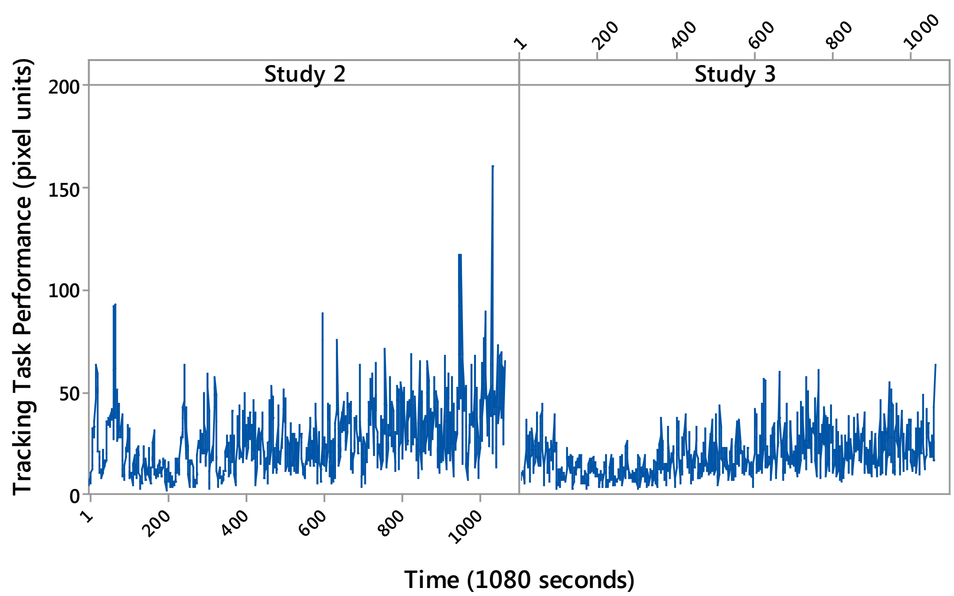


Figure C – 6 Tracking task performance: study 2 vs. study 3 (participant 6)

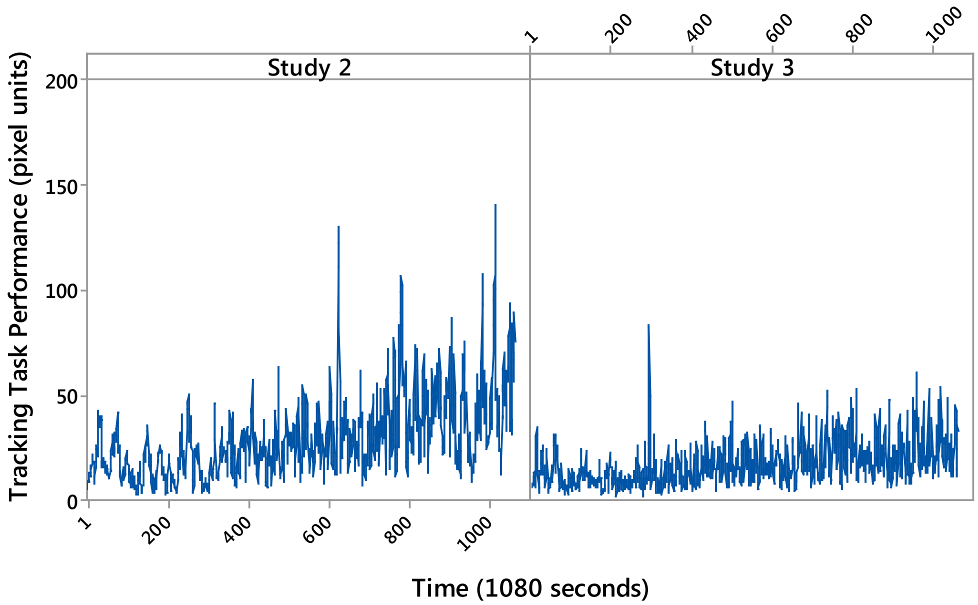


Figure C – 7 Tracking task performance: study 2 vs. study 3 (participant 7)

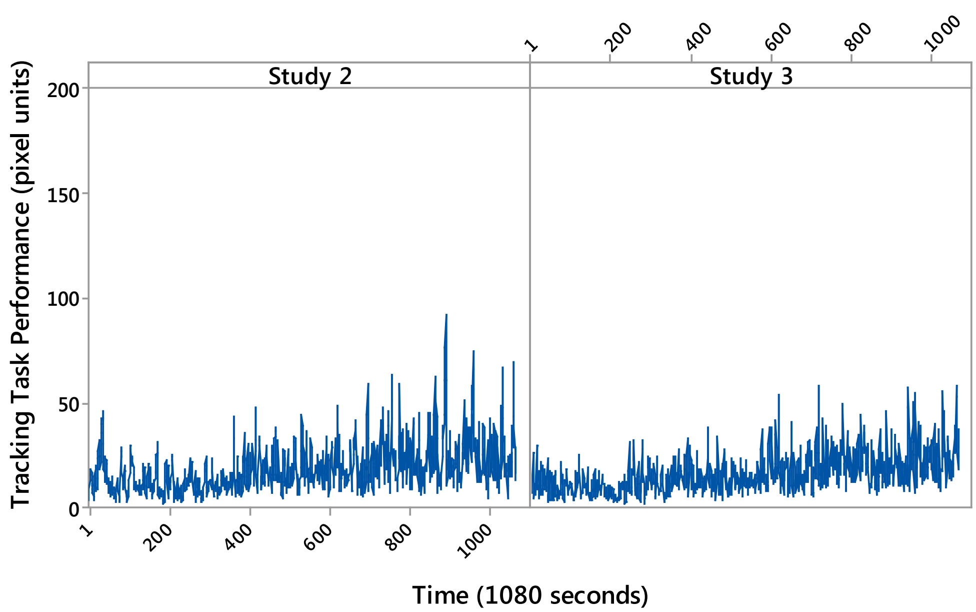


Figure C – 8 Tracking task performance: study 2 vs. study 3 (participant 8)

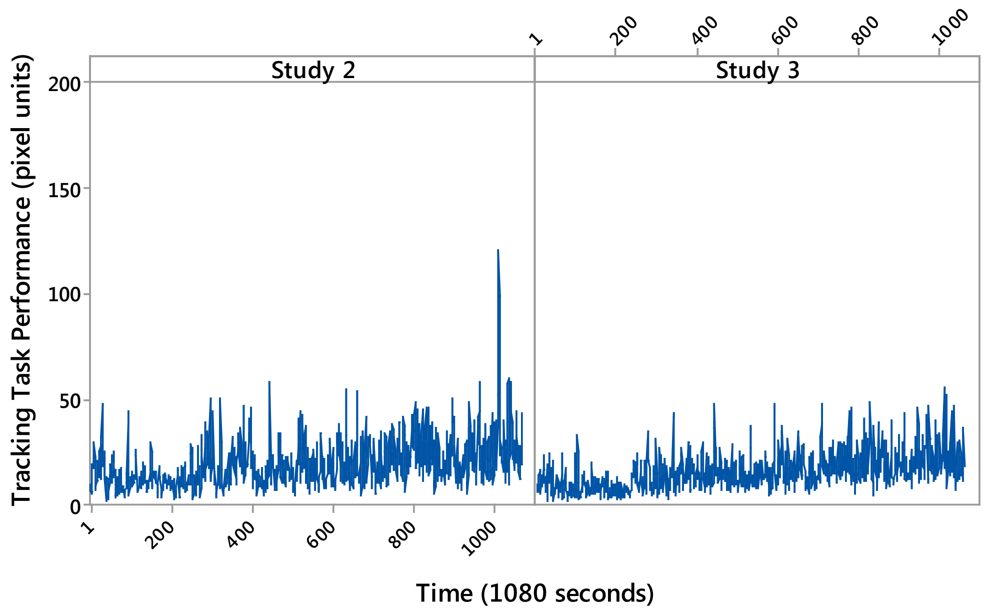


Figure C – 9 Tracking task performance: study 2 vs. study 3 (participant 9)

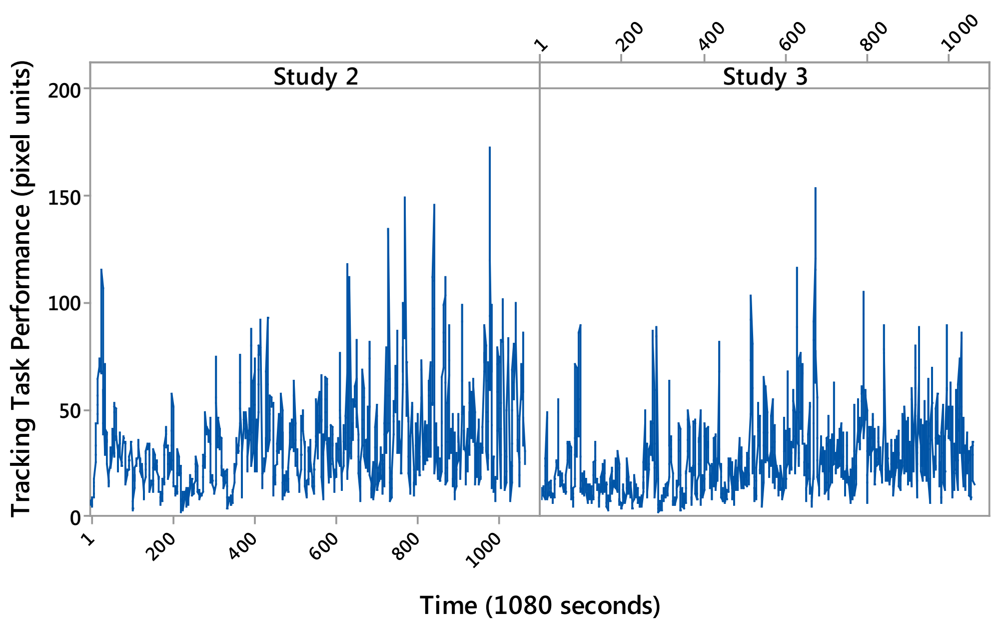


Figure C – 10 Tracking task performance: study 2 vs. study 3 (participant 10)

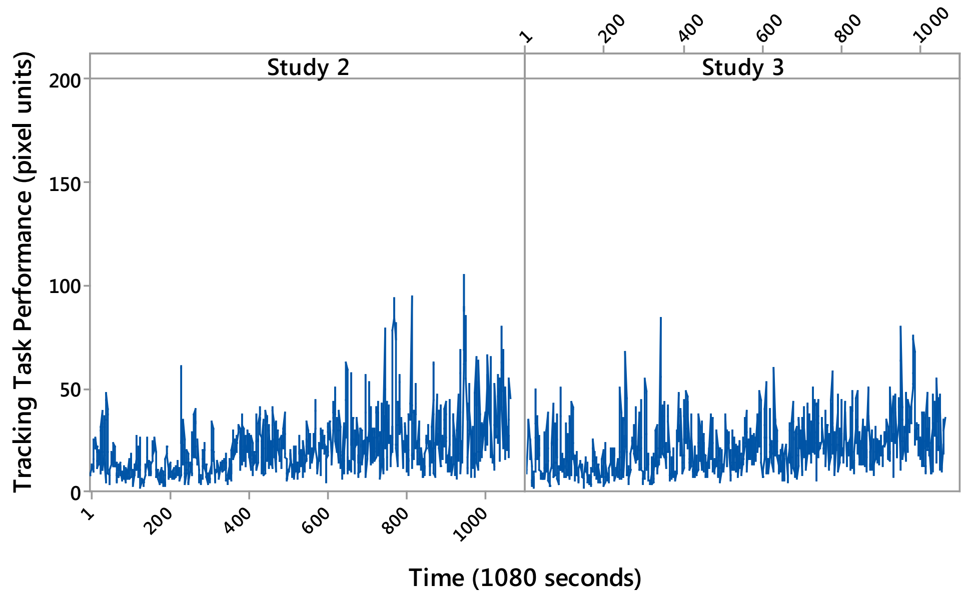


Figure C – 11 Tracking task performance: study 2 vs. study 3 (participant 11)

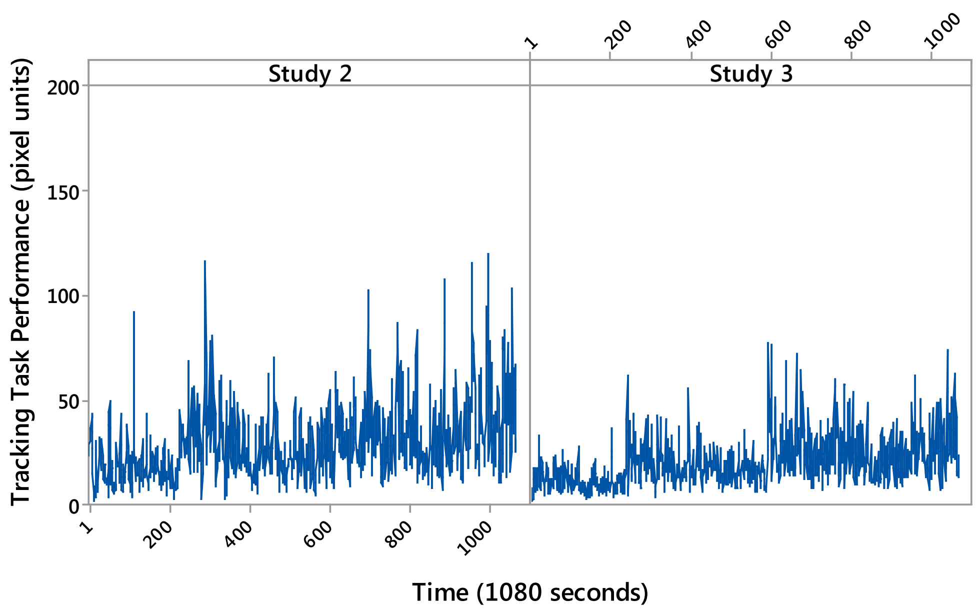


Figure C – 12 Tracking task performance: study 2 vs. study 3 (participant 12)

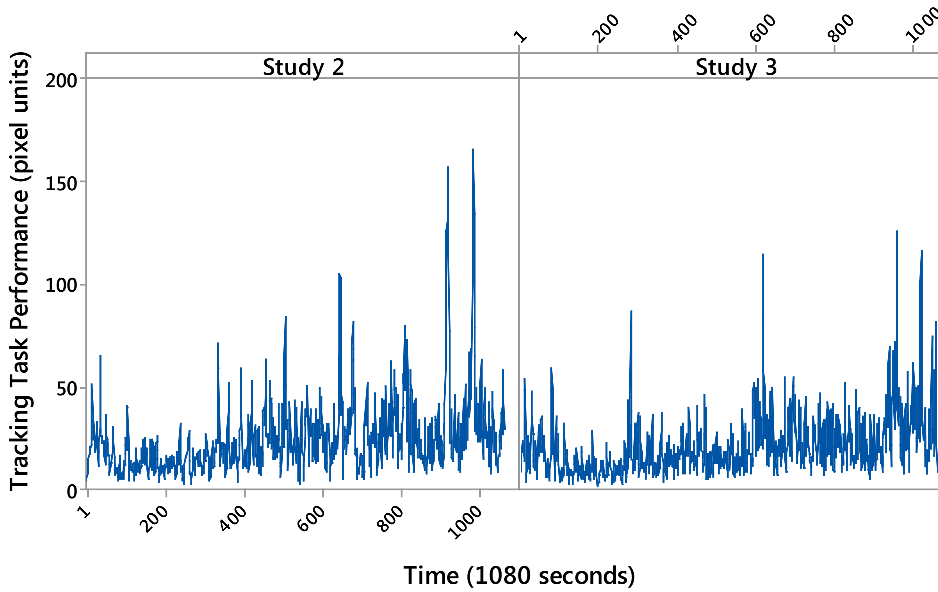


Figure C – 13 Tracking task performance: study 2 vs. study 3 (participant 13)

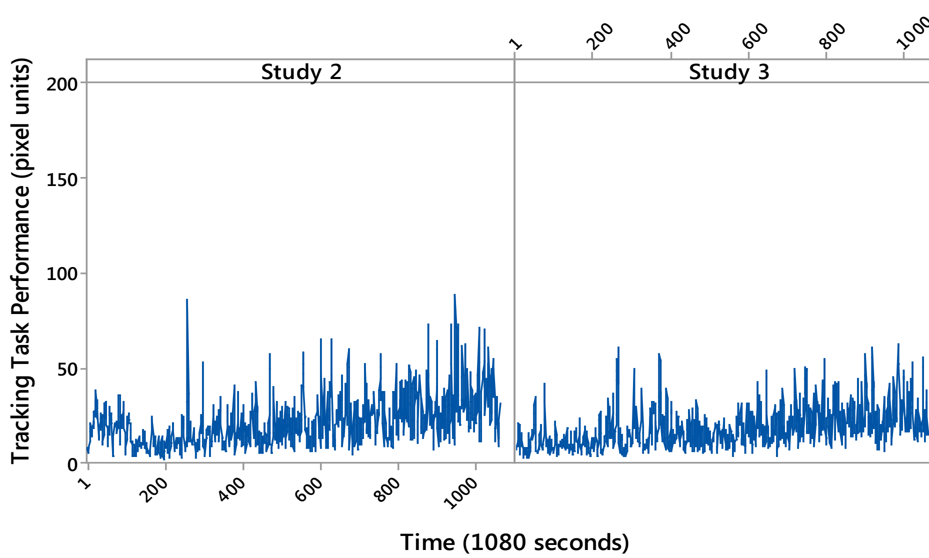


Figure C – 14 Tracking task performance: study 2 vs. study 3 (participant 14)

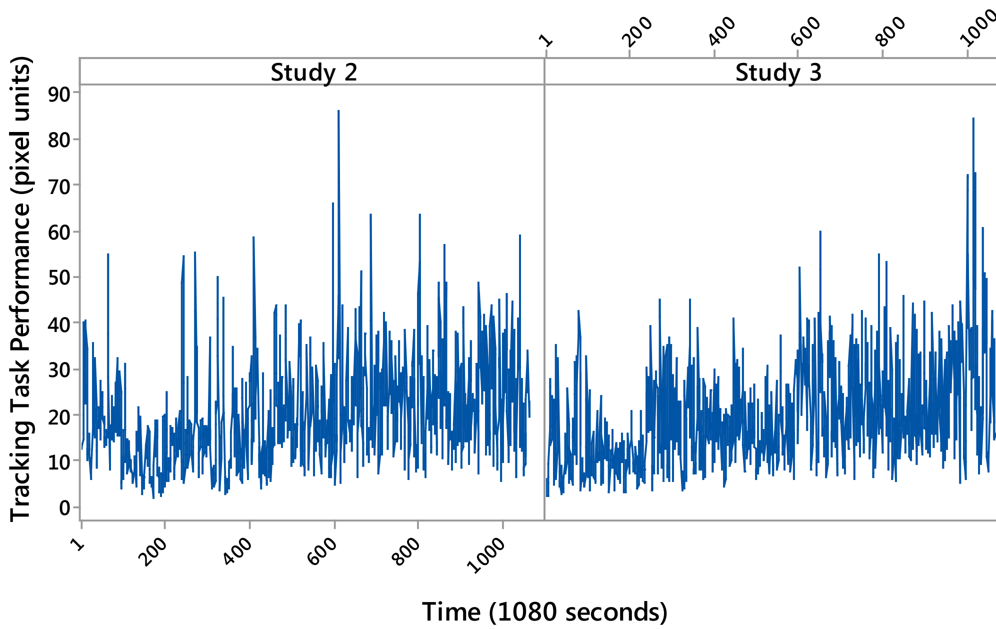


Figure C – 15 Tracking task performance: study 2 vs. study 3 (participant 15)

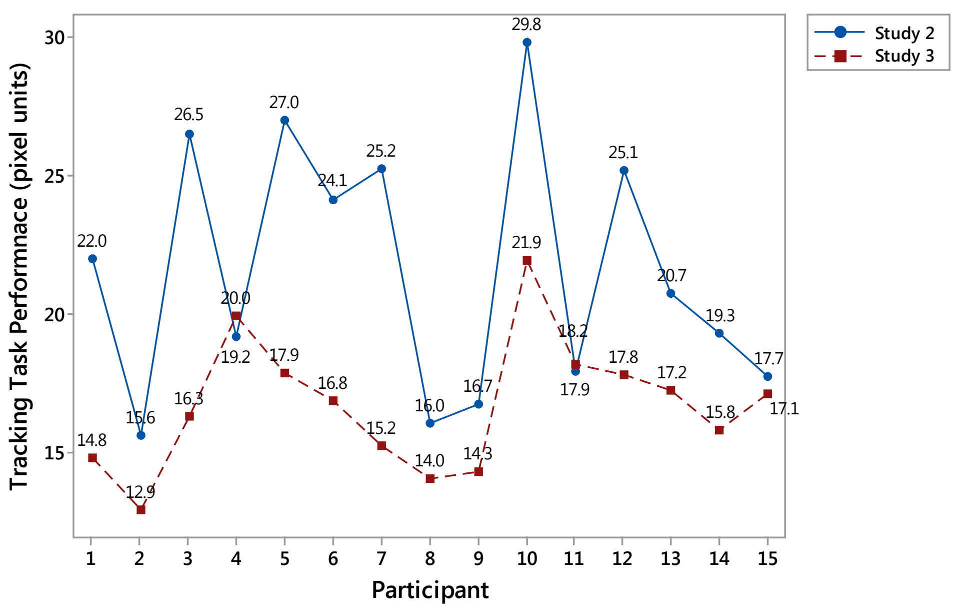


Figure C – 16 Tracking task performance (medians): study 2 vs. study 3

Appendix D STUDY THREE (DATA)

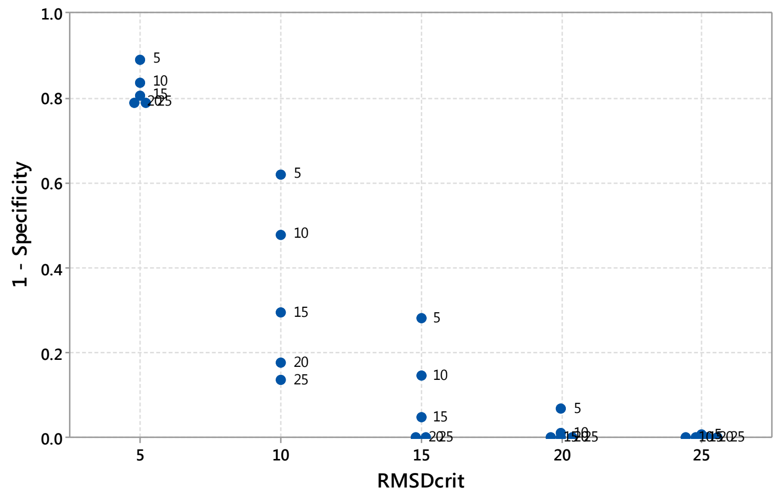


Figure D – 1 1- specificity: evaluation of parameters (RMSDcrit, ε) on participant 1's tracking task performance

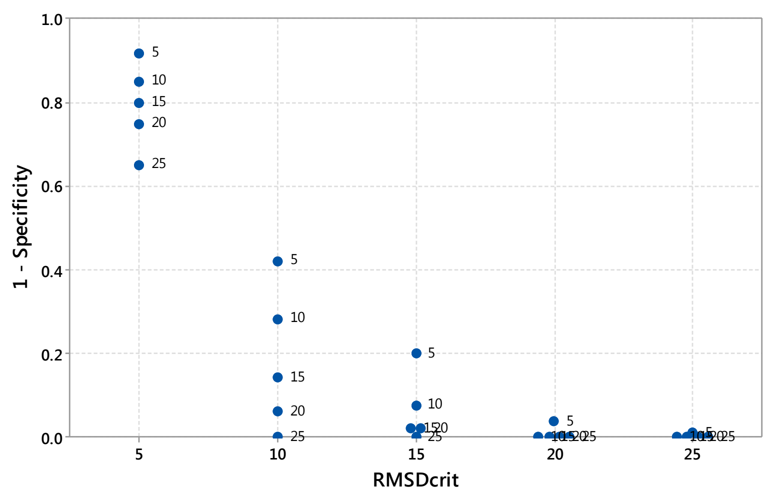


Figure D – 2 1- specificity: evaluation of parameters (RMSDcrit, ε) on participant 2's tracking task performance

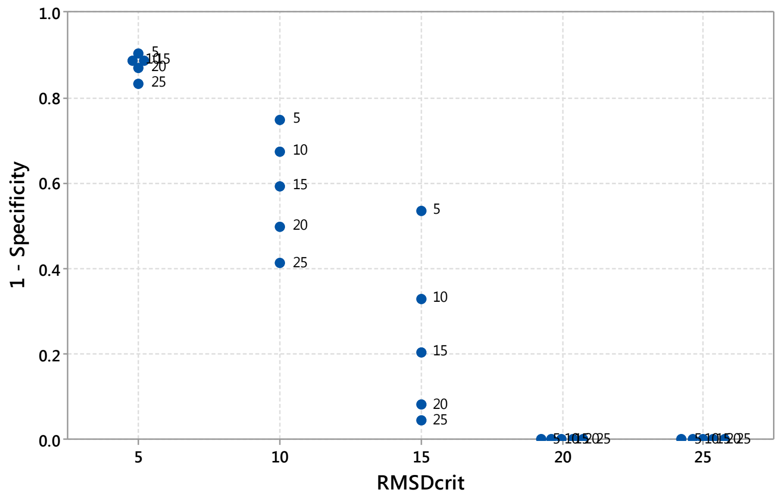


Figure D – 4 1- specificity: evaluation of parameters (RMSDcrit, ε) on participant 4's tracking task performance

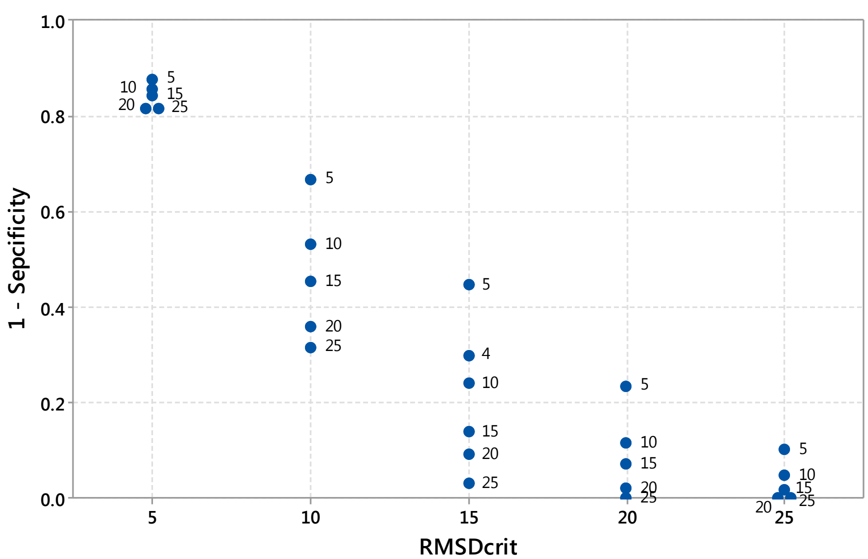


Figure D – 5 1- specificity: evaluation of parameters (RMSDcrit, ε) on participant 5's tracking task performance

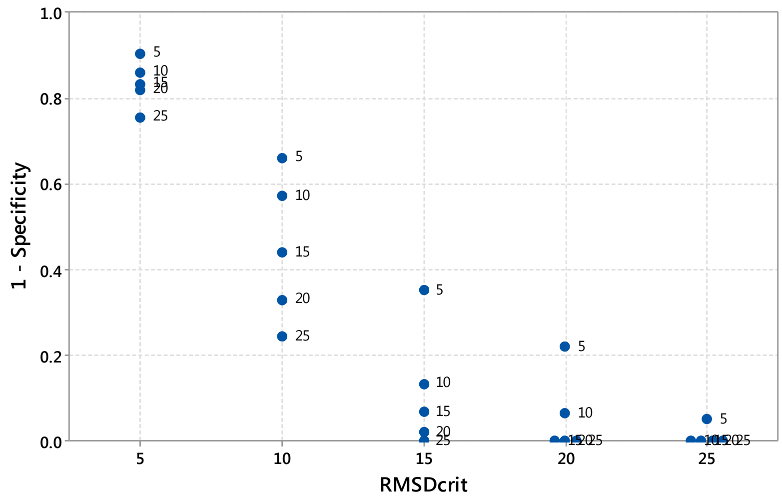


Figure D – 6 1- specificity: evaluation of parameters (RMSDcrit, ε) on participant 6’s tracking task performance

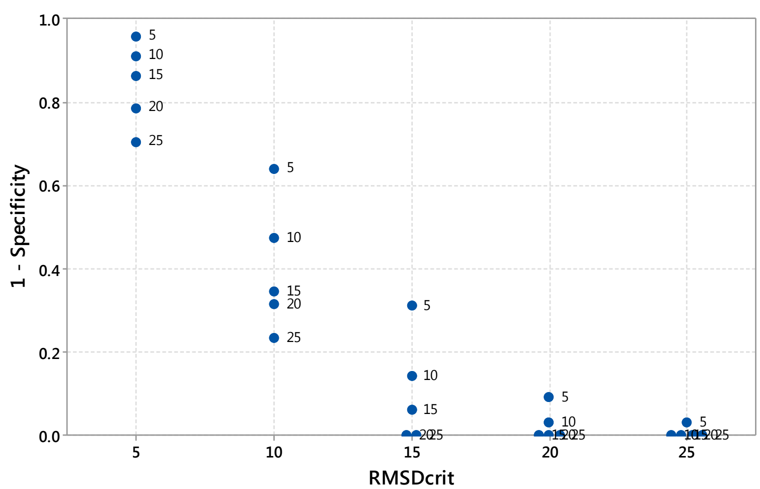


Figure D – 7 1- specificity: evaluation of parameters (RMSDcrit, ε) on participant 7’s tracking task performance

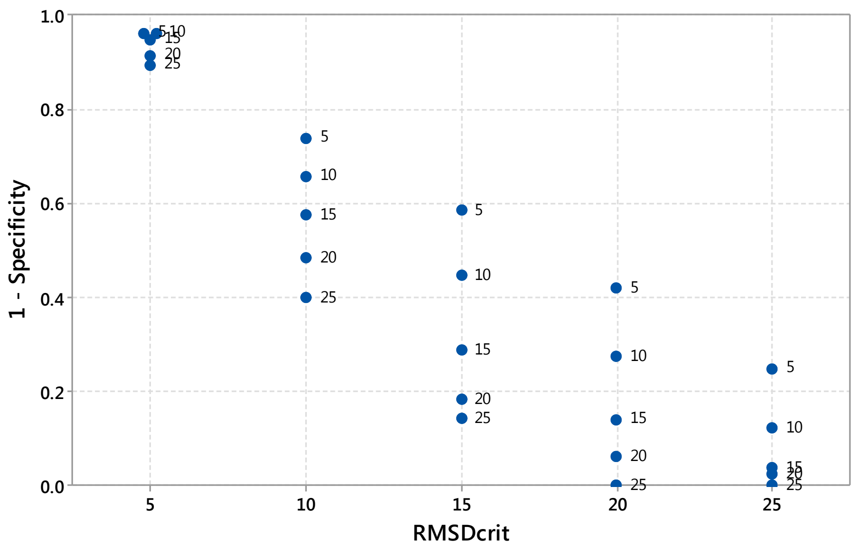


Figure D – 10 1- specificity: evaluation of parameters (RMSDcrit, ε) on participant 10's tracking task performance

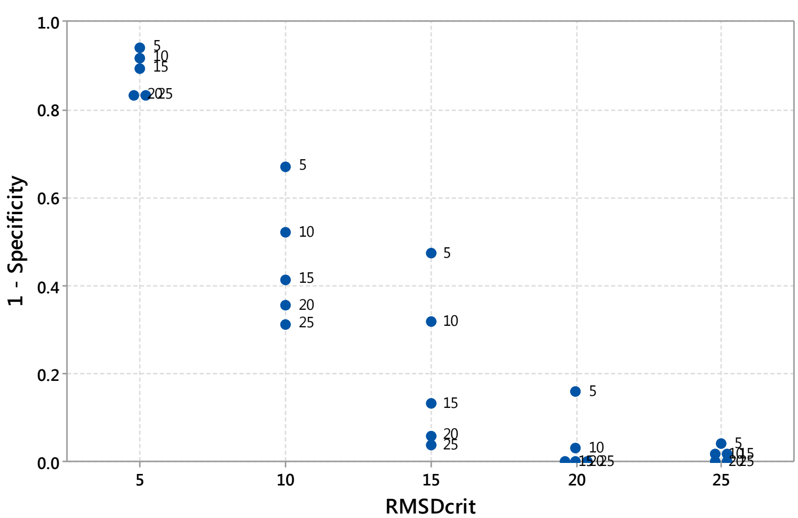


Figure D – 11 1- specificity: evaluation of parameters (RMSDcrit, ε) on participant 11's tracking task performance

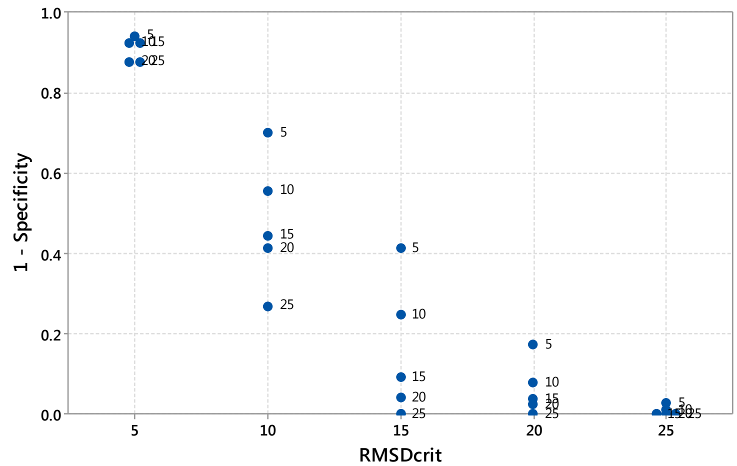


Figure D – 12 1- specificity: evaluation of parameters (RMSDcrit, ε) on participant 12's tracking task performance

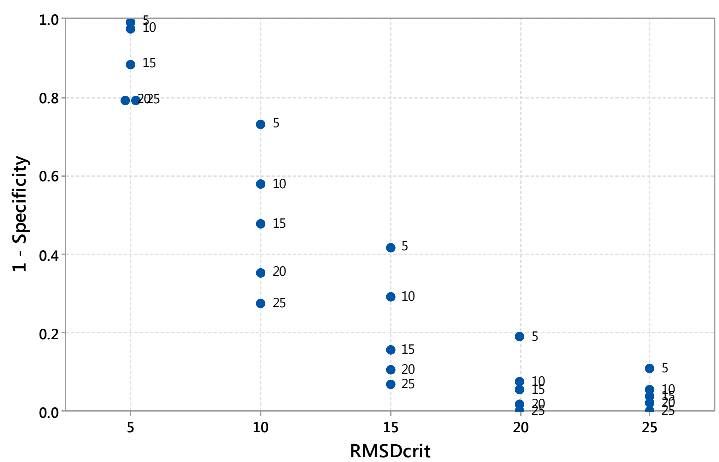


Figure D – 13 1- specificity: evaluation of parameters (RMSDcrit, ε) on participant 13's tracking task performance

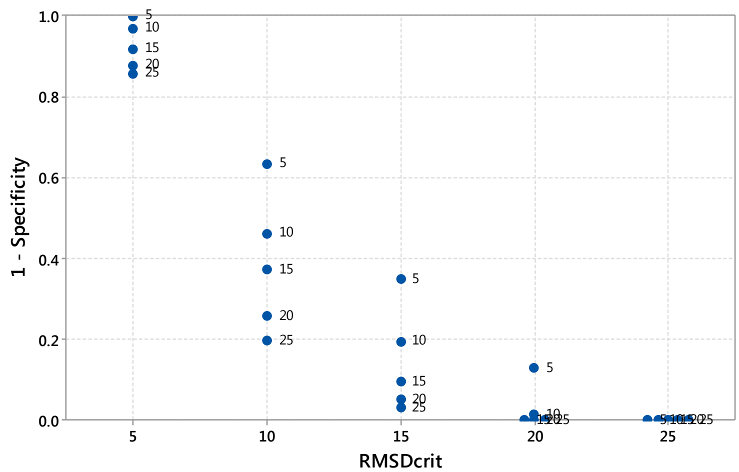


Figure D – 14 1- specificity: evaluation of parameters (RMSDcrit, ε) on participant 14’s tracking task performance

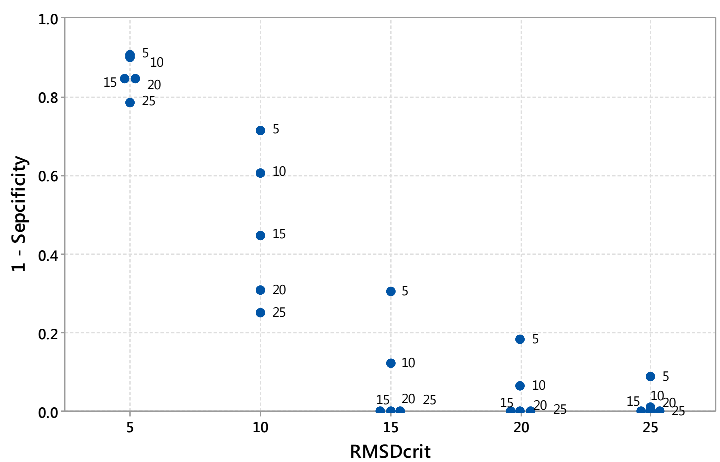


Figure D – 15 1- specificity: evaluation of parameters (RMSDcrit, ε) on participant 15’s tracking task performance

Appendix E APPROVED CONSENT FORM

For IRB Use Only

APPROVED
JAN 05 2015
EXPIRES 1-4-16
PURDUE UNIVERSITY
INSTITUTIONAL REVIEW BOARD

RESEARCH PARTICIPANT CONSENT FORM
Detection of Operator Performance Breakdown and its Use as a Trigger Mechanism for Function Allocation
Dr. Steven J. Landry
School of Industrial Engineering
Purdue University

What is the purpose of this study?

The purpose of this study is to identify the indications of the human performance breakdown (PB) to identify its use as a trigger mechanism for function allocation system. PB is an extreme condition where the operator no longer possesses control of the task(s). If such indications could be reliably detected, then it could be used as an automation triggering point to assist human operators. In order to participate in this study, you must be at least 18 years old (18 – 49).

What will I do if I choose to be in this study?

You will be participating in three studies. All three studies will be conducted on a same day. You have a full right not to participate in all three studies. In the first two studies, you will be asked to perform three tasks: system monitoring task, resource management task, and tracking task. These tasks, called MATB-II, are designed in a way that it mimics the tasks that pilots perform in the cockpit environment. The task difficulty of the tracking task will be manipulated to increase task demand, while the difficulty of the other two tasks will be maintained static. In the third study, you will be asked to perform only the tracking task.

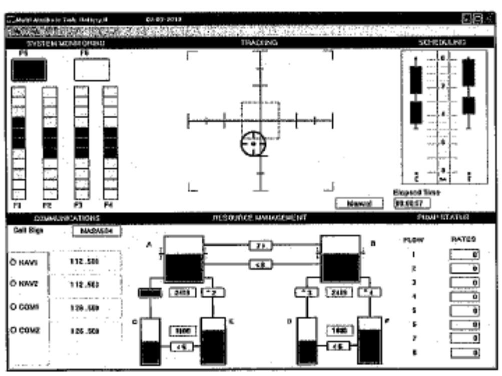


Figure 1. The secreenshot of MATB-II

In the tracking task, the target (blue dot with a circle) will randomly move. The goal is to control the position of the target with a joystick to keep it at the center point. For this task, the Root Mean Squared Deviation (RMSD) from the Center point is calculated to evaluate your performance.

In the resource management task, the goal is to control the volume of tank A and B within one or more of eight pumps. In the resource management task, fuel levels in two primary tanks (A & B) must be maintained at a target level (2,500 units) and violation of it will be considered an error.

The system monitoring task requires the operator to monitor and respond to simulated warning lights and gauges. You should respond when the green light goes off, or the red light comes on, or one of the scales to either shift to the top or bottom region. The minimum response time is set to be five seconds for all stimuli in this task. If you fail to respond within a given amount of time, then it will be counted as an error.

During your participation, heart rate, respiration rate, and your performance in the aforementioned three tasks in the MATB-II will be recorded. The physiological data will be gathered using the BioHarness™ 3 from Zephyr's BioHarness technology. This device will be placed on your chest using a chest strap, which incorporates Electrocardiography (ECG) and breathing detection sensors.

Figure 2 BioHarness™ 3

The overall process of activities is provided in appendix A. In the beginning of the first study, you will have sufficient time to familiarize yourself with performing the tasks. Then, you will be participating in actual experiment for approximately 25 minutes. You will have sufficient break period between each activity. In the beginning of the second study, you will have practice session to familiarize yourself with tasks, again. After the practice, your heart rate and breathing rate during the resting period will be collected for 2 minutes. Next, you will perform tasks for about 25 minutes. The third study will be conducted in a same way as the second study. But, you will be asked to perform only the tracking task.

<u>How long will I be in the study?</u>

You will be participating in three studies. Each study will last maximum one hour. Each study will be conducted on a same day. You have a full right not to participate in all three studies.

<u>What are the possible risks or discomforts?</u>

This experiment will present minimal risk, no more than everyday activities. Although the collected data on subjects will be safely locked in the room 4512 at Wang Hall, there is a potential for breach of confidentiality. The use of BioHarness™ 3 may cause itchiness as conductive ECG sensor pad has to touch skin. Also, it requires the participants to wear chest strap, which may cause uncomfortableness.

Are there any potential benefits?

There may be no direct benefit to you by participating in this experiment. However, the results of the study may help building more effective human automation interaction system. The finding could inform decisions regarding the appropriate allocation between humans and automation of functions associated with tasks of pilot and air traffic controller (e.g., separation assurance).

What happens if I become injured or ill because I took part in this study? (This section is only required if this study is greater than minimal risk.)

If you feel you have been injured due to participation in this study, please contact Dr. Steven J. Landry @ 765-494-6256, or email: slandry@purdue.edu. Purdue University will not provide medical treatment or financial compensation if you are injured or become ill as a result of participating in this research project. This does not waive any of your legal rights nor release any claim you might have based on negligence.

Will information about me and my participation be kept confidential?

The project's research records may be reviewed by San Jose State University Research Foundation and NASA Ames Research Center, and by departments at Purdue University responsible for regulatory and research oversight.

You will not be assigned with identifiers, and therefore data will not be labelled by a participant ID as well.

The collected data and this consent form will be locked in the room 4512 at Wang Hall at Purdue University to minimize the potential for breach of confidentiality. The Principal investigator (Steve Landry, PhD) will have immediate access to them for a minimum three years after the study is closed. They will be destroyed after three years.

What are my rights if I take part in this study?

Your participation in this study is voluntary. You may choose not to participate or, if you agree to participate, you can withdraw your participation at any time without penalty or loss of benefits to which you are otherwise entitled.

Who can I contact if I have questions about the study?

If you have questions, comments or concerns about this research project, you can talk to one of the researchers. Please contact Dr. Steven J. Landry @ 765-494-6256, or email: slandry@purdue.edu. If you have questions about your rights while taking part in the study or have concerns about the treatment of research participants, please call the Human Research Protection Program at (765) 494-5942, email (irb@purdue.edu)or write to:

Human Research Protection Program - Purdue University
Ernest C. Young Hall, Room 1032
155 S. Grant St.,
West Lafayette, IN 47907-2114

Documentation of Informed Consent

I have had the opportunity to read this consent form and have the research study explained. I have had the opportunity to ask questions about the research study, and my questions have been answered. I am prepared to participate in the research study described above. I will be offered a copy of this consent form after I sign it.

______________________________ ____________________
Participant's Signature Date

____
Participant's Name

______________________________ ____________________
Researcher's Signature Date

- The participant must sign and date the consent form. The only exception is if the study is granted a waiver of signed consent.
- The researcher's signature, above, refers to the research team member who has obtained the participant's consent. The researcher's signature indicates s/he has explained the research to the participant (or the legally authorized representative when IRB approved) and has answered any of the participant's questions

VITA

VITA

Hyo-Sang Yoo received his Bachelor's Degree in Industrial Engineering from Purdue University (December 2009). Hyo-Sang started working with Dr. Steven J. Landry at Purdue University as an Undergraduate Research Assistant in 2008. He joined the Landry's lab as a recipient of a Summer Undergraduate Research Fellowship (SURF) and continued working with Dr. Landry throughout the rest of his undergraduate years as part of his Honors program. He joined the Ph.D. program at the School of Industrial Engineering in the spring semester of 2010 and continued working with Dr. Landry on human factors problems in aviation. He has also been an active member of the Human Factors and Ergonomics Society (Purdue chapter) and Alpha Pi Mu (Industrial Engineering Honors Society). He joined the Airspace Operations Laboratory (AOL) in the Human Systems Integration Division at NASA Ames Research Center in May 2013. Since then, he has been involved in various projects related to NextGen air traffic management systems. After graduation, Hyo-Sang plans to accept an offer from AOL at NASA Ames Research Center.