ABSTRACT

LIU, SHIJING. Quantitative Modeling of User Performance in Multitasking Environments. (Under direction of Dr. Chang S. Nam.)

Multitasking is the human ability to engage in a variety of tasks simultaneously through task switching. This is one of the most important skills required for human operators to perform highly-complex and safety-critical jobs, such as air traffic controllers concurrently performing navigation, communication, and coordination, and commercial vehicle drivers talking on the phone and looking at the radio while driving. However, with the increase of the complexity of tasks, the required mental workload tends to increase and maintaining task performance within an acceptable level becomes more challenging. Individual differences in working memory capacity (WMC) have been estimated as predictors of varying cognitive abilities. The increased demand of cognitive resources may exceed the limitation of the human cognitive system, therefore leading to performance degradation and an increased occurrence of errors.

Previous studies have demonstrated different approaches for evaluating and improving user performance in a multitasking environment. Nevertheless, studies in the area of multitasking still present significant gaps: (1) few studies have quantitatively analyzed the information processing and user performance in multitasking, and there is a general lack of a quantitative approach to improve multitasking performance; and (2) the relationship between WMC, task difficulty, and user performance in multitasking still remains open.

The main goals of this study were to (1) propose and validate a quantitative model for user performance and improvement in a multitasking environment; and (2) investigate the relationship between WMC, task difficulty, and multitasking performance.

In this study, a quantitative model for user performance in multitasking was proposed. The Multi-Attribute Task Battery-II (MATB-II) was used in the experiments as a multitasking platform. The proposed model included quantification of stimuli from each MATB-II subtask as baud rate (bits per second), selection of task difficulty and task weight, as well as the rearrangement of task weights. This research followed a two-phase experimental approach.

The first phase consisted of two sessions: practice and MATB-II experiment sessions. The objectives of the first phase were to apply the proposed model and identify a performance baseline for each participant performing MATB-II tasks. Before the experiment, participants were asked to complete a span test to estimate their WMC. The automated operation span task (OSPAN) was applied in this study. The task involves a computerized span test requiring participants to remember items (letters) and solve math problems. At the beginning of Phase I experiment, all participants were required to complete a two-hour practice session in MATB-II. If their performance met the predefined criteria, they were selected to undertake the MATB-II experiment session. During the experiment session, selected participants completed a set of equally weighted MATB-II tasks with different levels of task difficulty. A performance baseline was calculated for each participant. Twenty-five participants were qualified and completed all sessions in Phase I.

The second phase consisted of three sessions: rearrangement of task weights, MATB-II experiment, and comparison. The objectives of the second phase were to validate the proposed model through a rearranged set of multitasks and performance comparison, and to investigate the relationship of WMC, task difficulty, and multitasking performance. All twenty-five participants from Phase I were invited to the second phase. During the second phase, a revised set of multitasks was designed for each participant according to their performance baseline and the proposed model. Then, the revised set of multitasks was tested by the select participants. Comparisons of user

performance were made between two phases to validate the proposed model. Results from the comparisons indicated improvement of user performance in the second phase. Statistical analyses were performed to investigate the effects of WMC and task difficulty. Results demonstrated the relationship between WMC, task difficulty and user performance that higher level of task difficulty led to decreased performance and participants with high WMC had an overall high level of performance.

The proposed model attempted to offer an approach to quantify the information from machines and the response from human operators in a multitasking environment. From a practical point of view, the current research may make significant contributions to improve the design of multitasking systems and training procedures for human operators. From a theoretical perspective, this research provides a framework to quantitatively evaluate multitasking systems and human performance in order to understand the interaction between systems and human operators.

Quantitative Modeling of User Performance in Multitasking Environments

by
Shijing Liu

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Industrial Engineering

Raleigh, North Carolina
2017

APPROVED BY:

Chang S. Nam
Committee Chair

David A. Dickey

David B. Kaber

Robert St. Amant

Nancy Currie

DEDICATION

To my parents.

BIOGRAPHY

Shijing Liu is a PhD candidate in the Edward P. Fitts Department of Industrial and Systems Engineering at North Carolina State University, with a Bachelor of Industrial Engineering from China University of Mining & Technology, Beijing, and a Master of Industrial and Systems Engineering from Ohio University.

ACKNOWLEDGMENTS

I would like to express my sincere appreciation to Dr. Chang S. Nam. His continuous guidance and instruction kept me on track throughout my doctoral program. Without his expertise and insight, I would not have been able to complete all my research and develop both personally and professionally.

I would also like to acknowledge each member of my committee, Dr. David Dickey, Dr. David Kaber, Dr. Nancy Currie, and Dr. Robert St. Amant for their valuable feedback and suggestions. Additionally, I would like to thank Dr. Mohamed Bourham for serving as a Graduate School Representative for my dissertation defense. Further, I would like to give my special thanks to Dr. Shu-Cherng Fang, who inspired me and provided lots of recommendations on my dissertation.

Acknowledgements are also made for the help and support rendered by all my research fellows at Ergonomics Lab, especially for Amy Wadeson, Na Young Kim, and Kyle Bond who helped me a lot in the pilot test for my research. Special thanks go to Ellen Wittenberg, who proofread this manuscript with incredible patience.

Finally, I am sincerely thankful to my parents and fiancé, Yunlong, for their unwavering love and encouragement along the way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF CONTRIBUTING PUBLICATIONS

Liu, S., Wadeson, A., & Nam, C. S. (2016, July). Toward Quantitative Modeling of User Performance in Multitasking Environments. In *International Conference on Engineering Psychology and Cognitive Ergonomic*s (pp. 3-9). Springer International Publishing.

Liu, S., Wadeson, A., Kim, N. Y., & Nam, C. S. (2016, September). Effects of Working Memory Capacity, Task Switching, and Task Difficulty on Multitasking Performance. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 60, No. 1, pp. 502-506). SAGE Publications.

LIST OF ACRONYMS

MATB	Multi-Attribute Task Battery
AF_MATB	Air Force MATB
ANOVA	Analysis of Variance
bps	bit per second
BKD	Blink duration
BKR	Blink rate
EF	effort
FR	frustration
HCI	Human-Computer Interaction
MD	mental demand
NASA-TLX	NASA Task Load Index
OSPAN	operation span
PD	physical demand
PE	performance
RMSD	root-mean-square deviation
RMSE	root-mean-square error
RR	response ratio
RT	response time
TD	temporal demand
WMC	working memory capacity
WRS	Workload Rating Scale

1 Introduction

Performing more than one task at a time in highly-complex and safety-critical jobs is increasingly common in modern life, such as air traffic controllers concurrently performing navigation, communication, and coordination, or commercial vehicle drivers talking on the phone and looking at the radio while driving (Colom, Martínez-Molina, Shih, & Santacreu, 2010; König, Bühner, & Mürling, 2005; Lin, 2013). In general, multitasking is defined as the human ability to engage in a variety of tasks simultaneously through task switching (Delbridge, 2000; Hambrick, Oswald, Darowski, Rench, & Brou, 2010; Junco & Cotton, 2012).

Due to the increasing complexity of our modern, technology-driven society, people are increasingly engaged in multitasking and there are many advantages to this. For instance, completing tasks simultaneously can save time and increase productivity (Colom et al., 2010; Spink, Park, Jansen, & Pedersen, 2006). However, when tasks increase in complexity, the required mental workload tends to increase and maintaining task performance within an acceptable level becomes more challenging (Aricò, Borghini, Graziani et al., 2014). The increased demand of cognitive resources may exceed the limitation of the human cognitive system, therefore leading to performance degradation and an increase in errors (Aricò et al., 2014; Du & Spink, 2010). However, it has been shown that effective training can greatly reduce such limitations and increase multitasking performance (Dux, Tombu, Harrison et al., 2009; Erickson, Colcombe, Wadhwa et al., 2007; Schneider, 1985; Schneider & Detweiler, 1988).

Working memory is responsible for certain cognitive functions in multitasking, such as temporary storage and manipulation of the information necessary for a wide range of complex cognitive activities (Baddeley & Hitch, 1974; Colom et al., 2010; König et al., 2005). Increasing

evidence has shown that individual differences in working memory capacity (WMC) were important predictors of various cognitive abilities (Kane & Engle, 2003; Redick & Engle, 2006; Engle & Kane, 2004; Unsworth, Heitz, Schrock, & Engle, 2005).

Previous studies have demonstrated different approaches for evaluating and improving user performance in a multitasking environment. Individual differences in WMC have also been investigated in multitasking environments. However, studies in the area of multitasking still present significant gaps: (1) few studies have quantitatively analyzed the information processing and user performance in multitasking, and there is a general lack of a quantitative approach to improve multitasking performance; and (2) the relationship between WMC, task difficulty, and user performance in multitasking still remains open.

To overcome these limitations, this study was intended to (1) propose and validate a quantitative model for user performance and improvement in a multitasking environment; and (2) investigate the relationships between WMC, task difficulty, and multitasking performance.

The following sections introduced cognitive modeling in multitasking performance and improvement, the individual differences of WMC in multitasking, the limitations of previous studies, and the motivation and framework of this study.

1.1 Modeling of multitasking performance

1.1.1 Multitasking and Cognitive Models

To evaluate and predict user performance in a multitasking environment, various types of cognitive models and approaches have been proposed in previous research. The studies of multitasking usually examined how well each task would be performed in a multiple task combination, in comparison with how each task would be performed alone (Wickens, Hollands, Banbury, & Parasuraman, 2015). These approaches can be separated into two categories: the first in which researchers focus on assessing the special human ability for multitasking, sometimes referred to as the "ability of time sharing" (Brookings, 1990; Hambrick et al., 2010; Horrey & Wickens, 2004); the second in which the researchers focus more broadly on multitasking as a class of situations (Hambrick et al., 2010; Spink, Cole, & Waller, 2008).

For the approaches assessing the ability of time sharing, Wickens, Dixon, & Ambinder (2006) described a computational model to predict total interference between a time-shared pair of tasks. Two characteristics that account for variance in time-sharing efficiency were established (Wickens, 2008): (a) "the extent to which time-shared tasks used the same versus different processing structures" (p. 450) (e.g., auditory versus visual processing (Guastello et al., 2013)), and (b) "the extent to which 'difficulty insensitivity' was expressed when the two tasks used different structures" (p. 450). "Difficulty insensitivity" refers the phenomenon that degradation of the performance is not observed from one task when the difficulty increases in a concurrent task (Wickens, 1991, 2008). Therefore, some combinations of multitasks are performed better together than other combinations (Guastello et al., 2013; Hambrick et al., 2010).

Some approaches defined multitasking as situations in which the human operator must make conscious shifts of attention between two or more tasks (Spink et al., 2008). In a study of the determinants of multitasking (Oswald, Hambrick, & Jones, 2007), a model for investigating potential predictor variables on a continuum from distal (indirect) to proximal (direct) causes was provided. Distal predictors include general dimensions of psychological functioning, including intelligence and personality, while proximal predictors include task-specific knowledge and strategies, and state variables like goal orientation, anxiety and perceived workload. A web search model (Du & Spink, 2011), which integrated multitasking, cognitive coordination, and cognitive shifts during multitasking, was presented to explore the characteristics of multitasking behavior, types of cognitive shifts, and levels of cognitive coordination as well as the relationships between them during a web search. Alexopoulou, Morris, and Hepworth (2014) also presented an integrated model for multitasking during web searching. According to this model, they investigated how information behavior is affected by working memory, cognitive coordination, cognitive shifts, and various artifacts and task variables influenced by the PAT model (Personal, Artifact and Task characteristics) of flow.

Meanwhile, various cognitive models have quantitatively analyzed human performance in single or multitasking scenarios. Many models discussed the quantification of information in a system with the human operator as a transmitter of information, for instance, the applications of Hick-Hyman law (Hick 1952; Hyman 1953) and Fitts law (Fitts, 1954), either singly or in combination (e.g., Clark & Ivry, 2010; Corbett, Yamaguchi, Liu et al., 2013; Phillips, Repperger, Kinsler et al., 2007; Seow, 2005). These models are originally derived from Shannon's Theory, a fundamental theorem of communication systems (Shannon, 1948). A major achievement of these cognitive models of human-computer interaction is the quantification of information transmitted

from the system as a baud rate (bits per second) (Crossman, 1960; Repperger, Phillips, & Chelette, 1995).

These cognitive models paved the way for future researchers to interpret the relationship between human performance and multitasking. However, it is still unclear how to improve human performance in multitasking through a quantitative approach.

1.1.2 Task Difficulty and Mental Workload

The effect of task difficulty has been investigated in both single and multitask scenarios. This effect is fairly straightforward and easy to understand in terms of the relationship between resources and performance (Wickens, 1991): when an increase in task difficulty occurs in a system, more resources from the human operator will be required to maintain performance at an optimal level. In this case, resources mean "the mental effort that is invested to improve performance" (Wickens, 1991).

During multitasking performance, mental effort is used to describe the amount of information processing resources. Previous studies have verified that a greater mental effort is required to maintain performance as task difficulty increases if the tasks are resource limited (Backs & Seljos, 1994; Durkee, Geyer, Pappada, Ortiz, & Galster, 2013). The manipulation of task difficulty has been an effective method for inducing varying levels of operator workload (Durkee et al., 2013; Fournier, Wilson, & Swain, 1999), and has been broadly applied in studies on workload and multitasking performance.

In a study of workload modeling in a multitasking environment (e.g., MATB-II), the difficulty of an ongoing task represented the most validated and successful parameter of the model,

reflecting the overall cost of limited cognitive resources and its effect on decision making (Gutzwiller, Wickens, & Clegg, 2014). Another study on the neurophysiologic monitoring of mental workload in multitasking demonstrated the manipulation of task difficulty and its effect (Smith & Gevins, 2005). They manipulated task difficulty in MATB to increase working memory and decision making demands, motor control demands, and visuoperceptual demands. Then, significant changes which were accompanied with task difficulty were found in cortical activation over frontal, central, and posterior regions of the scalp.

Though task difficulty has been selected as an effective method to manipulate mental effort and decision making demands, there are limited instructions pertaining to the manipulation of task difficulty using a quantitative model and referring to the information processing resources.

1.1.3 Multitasking Strategy

To multitask, the human operator needs to shift attention to perform several independent but concurrent tasks. Benbunan-Fich, Adler, and Mavlanova (2011) implied that there are some temporal overlaps in a specific period of time during multitasking. To overcome these temporal overlaps, different task strategies have been evaluated and compared in previous studies. There are two types of task strategies that have been discussed and evaluated in much of the research on multitasking performance: serial and parallel strategies (Benbunan-Fich et al., 2011; Brumby, Salvucci, & Howes, 2009; Kaber & Kim, 2011; Zhang & Hornof, 2014).

In a serial strategy, operator completes each task in sequence and each task starts after the completion of the previous one (Bluedorn, Kaufman, & Lane, 1992). Although there is no temporal overlap between tasks, Benbunan-Fich et al. (2011) implied that the serial approach can be useful

to establish a baseline for comparing user performance with different types of multitasking approaches.

In a parallel strategy, operator needs to handle all concurrent tasks at the same time (Bluedorn et al., 1992). Therefore, there is a maximum degree of temporal overlap of these tasks. Previous study argued that it is difficult to achieve true parallel performance in practice, since human attention cannot be divided among many tasks simultaneously (Salvucci & Taatgen, 2011).

Previous studies on cognitive modeling also reveal that the application of different multitasking strategies will affect performance and impact safety (Brumby, Howes, & Salvucci, 2007; Brumby et al., 2009; Kaber & Kim, 2011; Zhang & Hornof, 2014). Comparison between these two types of task strategies has been explored in some studies on multitasking.

An example is found in Kaber & Kim (2011), in which the differences between parallel and serial strategies were evaluated through adaptive automation dual-task performances and the application of computational cognitive models (Kaber & Kim, 2011). They tested the effects of advanced auditory cuing of control mode changes in an adaptive automated system (Ballas et al., 1992) on human performance and explained the cognitive behaviors at mode changes by using a computational cognitive model. In their initial model, a computational GOMS (goal, operators, methods, and selection) language model with a serial (lockout) task processing strategy was applied. The revised model used a parallel task processing strategy and the results showed that user performance (response time and RMSD (root mean square deviation) in tracking task) improved.

Similarly, many studies revealed that human performance can be better predicted by a parallel strategy or that a human operator can perform at a higher level under a parallel strategy

with limited resources (Brumby et al., 2007, 2009; Fischer & Plessow, 2015; Sigman & Dehaene, 2008). Therefore, a parallel task strategy was applied in this study to better construct the model and better instruct users in system operation.

1.1.4 Training in Multitasking

Practice increases skills and abilities. Previous studies investigating the effects of practice on a single task and multitasking showed that effective training (e.g., prolonged and extensive) can greatly reduce multitasking interference (Dux et al., 2009; Ruthruff, Johnston, & Selst, 2001 & 2013).

Damos (1991) suggested that the dual-task performance should achieve some stable level before the data are collected. Similarly, in a multitasking environment, it is expected that operators need to practice until the achievement of a stable or asymptotic performance. Two types of human operator practice techniques combining single and multiple tasks for the purposes of achieving stable or asymptotic performance have been explored in previous studies (Damos, 1991).

The first practice technique utilized in the research examines the effect of varying amounts of single-task practice on multitasking performance. In these studies, two or more types of tasks were selected, but the training only focused on single task for each training session in order to assess the effect of the amount of single-task practice on multitasking performance. Task combinations included a reaction time task and a mental arithmetic task (Bahrick, Noble, & Fitts, 1954), a spatial task and a mental arithmetic task (Damos, 1986), a tracking task and several decision making tasks (Bowers, Christensen, & Eggemeier, 2014), etc. After the single task training session, participants were required to complete a combination of these tasks in the experimental session. The results from these experiments employing this type of practice

combination showed little effect of the amount of single-task practice on subsequent multitasking performance.

The second type of practice technique is used to test the effect of varying the amount of single and multitasking practice on multitasking performance. These types of studies used similar task combinations to the first type (e.g., tracking task, decision making task, mental rotation task, etc.). During the training session, participants focused on practicing the combination of these single tasks. The results implied that multitasking practice can produce better subsequent multitasking performances than the first type of practice technique (Damos, 1991; Folds, Gerth, & Engelman, 1987; Reick, Ogden, & Anderson, 1980).

Hence, to investigate user performance in multitasking, a stable or asymptotic performance should be achieved before experimental data are collected.

1.1.5 Individual Differences in Multitasking

The topic of individual differences has been widely addressed in research on user performance and behavior. With the increase of complex automated systems, exploration of individual differences in multitasking performance may provide insight into personnel selection (Salomon, Ferraro, Petros et al., 2015). Specific characteristics have frequently been used as criteria for selecting individuals who may be well suited for a particular task or job (Hurtz & Donovan, 2000).

Research on personality measures in multitasking usually include attention (Kane & Engle, 2003), intelligence (Colom et al., 2010; Salomon et al., 2015), working memory capacity (WMC) (Colom et al., 2010; Kane & Engle, 2003; Liu, Wadeson, Kim, & Nam, 2016; Sörqvist, 2010),

processing speed (Hambrick et al., 2010), the ability of task switching (Liu et al., 2016; Sanchez-Cubillo, Perianez, Adrover-Roig et al., 2009), timesharing ability (Brookings, 1990; Wickens, Mountford, & Schreiner, 1981), and many other aspects of individual differences.

In addition to these individual characteristics, WMC has also been examined in multitasking research as an important predictor of varying cognitive abilities (Kane & Engle, 2003; Engle & Kane, 2004). It has been shown to predict performance in various cognitive tasks, e.g., following directions, writing, note-taking, computer language learning, and bridge playing (Engle & Kane, 2004; Redick & Engle, 2006; Unsworth et al., 2005). But only a limited amount of research has explored the relationship between individual differences in WMC and multitasking performance through a quantitative model.

1.2 Limitations of Current Research on Multitasking Performance

As discussed in the preceding sections, various cognitive models and approaches have been demonstrated to assess user performance in multitasking. However, only a limited number of the proposed approaches provided quantitative methods to evaluate the information content transmitted between systems and human operator. Even fewer studies have shown how to improve user performance through a quantitative method in terms of human information processing in a multitasking environment. Therefore, it is necessary to develop an approach to quantitatively analyze multitasking performance and to provide a quantitative model to improve user performance in multitasking.

With the increase in complexities of modern technologies, proper training procedures are necessary for many highly-complex and safety-critical professions. While the effect of practice in multitasking has been evaluated in a wide range of tasks, from single to dual to multitasks, there is still a lack of detailed instructions and user-specific training procedures for complex multitasking.

Task difficulty is an effective method to manipulate mental demand in multitasking. And individual differences in WMC are important predictors in the research on multitasking. To assess multitasking performance, the effects of task difficulty and individual differences in WMC have been evaluated in previous studies. However, a limited amount of studies have considered task difficulty and individual differences in WMC through a quantitative model in multitasking environments. Thus, more research is still required to examine the relationships between WMC, task difficulty, and user performance in multitasking.

1.3 Research Motivation

The limitations of previous studies provided research directions for the development of the present study. It was necessary to provide an approach to quantitatively analyze multitasking

performance and the effects of task difficulty and individual differences in WMC in multitasking. Therefore, the objectives of this research were:

a) Propose and validate a model to quantify the information transmitted and processed between human operator and system, and provide the approach to evaluate and improve human performance in a multitasking environment.

b) Investigate the relationship between WMC, task difficulty, and human performance in a multitasking environment.

To achieve these objectives, a quantitative model for multitasking performance was proposed and a two-phase study was conducted. The framework of this research is shown in Figure 1.1.

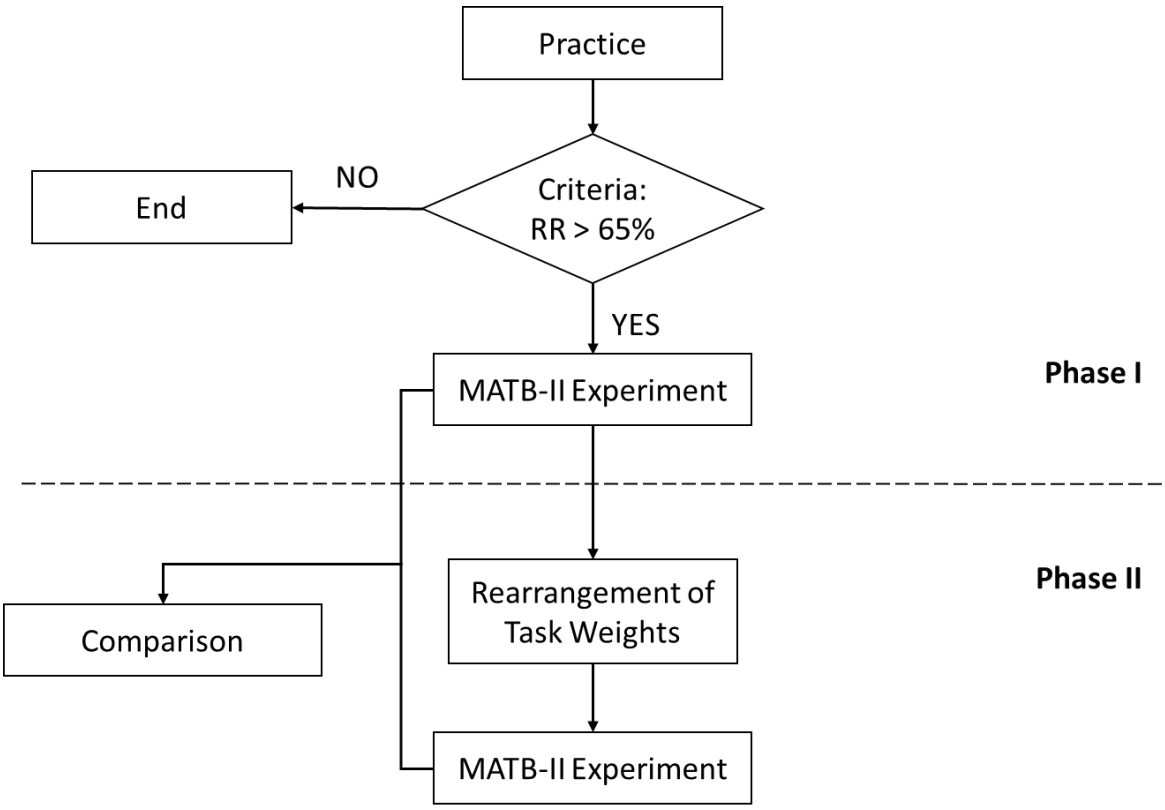


Figure 1.1 Research Framework

Note: RR represents "response ratio (%)," which is the ratio of the number of correct responses of subtask in a trial to the total number of events of subtask in a trial.

In the preliminary phase of this study, the proposed model was used in a multitasking environment, MATB-II, and a performance baseline for each participant was identified. To investigate the effects of individual differences in WMC, a screening session was performed during the recruitment of participants at the beginning of the first phase. The practice session trained and selected qualified participants based on a set of predefined performance criteria. Then, the MATB-II experiment session was completed by selected participants. The results from the first phase provided a baseline performance for the second phase.

During the second phase, the rearrangement of task weights for each individual was performed according to the proposed model. A MATB-II experiment session was completed by selected participants from the first phase. Comparisons of user performance were made between the experimental results from two phases.

In Chapter 2, a literature review was presented covering four relevant issues for this research, including multitasking and user performance, user performance modeling, and individual differences and practice in multitasking.

Based on the knowledge obtained from the related studies, research hypotheses and a detailed research plan was introduced in Chapter 3.

Chapter 4 proposed a quantitative model for user performance in a multitasking environment. The MATB-II was employed as a multitasking platform in this study and the proposed model is based on the application of MATB-II tasks. In the proposed model, information content from the multitasking system was quantified as baud rate (bits per second). The approach includes the baud rate selection and task difficulty for each individual task in MATB-II, task weight and overall baud rate selection, and the rearrangement of task weights.

Chapter 5 and 6 explained the details of the experiment methodology for Phase I and Phase II studies. Chapter 7 showed and discussed the experiment results. Chapter 8 concluded with a summary of findings and limitations of this study and explained research applications and future work.

2 Literature Review

2.1 Multitasking and User Performance

This section reviews a range of studies focused on multitasking and user performance, including the definitions of multitasking, user performance, and multitasking platforms.

2.1.1 Definitions of Multitasking

Multitasking is a frequently used term to describe the situation in which an operator needs to perform several tasks during a specified time period and it is a topic of increasing importance (Delbridge, 2000; Hambrick et al., 2010). But there is no universally accepted definition of multitasking. Does multitasking require an operator to engage in various activities simultaneously or engage in multiple tasks sequentially? Does it mean engagement in a single task while passively processing a secondary source of stimulation or input? Does the operator need to switch between different tasks or process all subtasks at the same time? If the operator needs to switch between tasks, are these tasks discrete or continuous? These questions imply the need for a clear definition of multitasking, including multitasking strategy, task type, and task operation. A summary of the definitions of multitasking in previous studies is shown in Table 2.1.

Table 2.1 Definition of Multitasking

Task Strategy	Task Type	Task Operation	Representative Study
Sequential	multiple tasks	switch between tasks	Delbridge (2000)
	multiple tasks	switch between tasks	Dzubak (2008)
	multiple tasks	shift between tasks	Spink et al. (2008)
	multiple tasks	switch between tasks	Oswald et al. (2007)
	dual tasks	switch between tasks	Logan & Gordon (2001)
	dual tasks	switch between tasks	Monsell (2003)
Parallel	multiple tasks	switch between tasks	Colom et al. (2010)
	main and secondary tasks	switch between tasks	Junco & Cotton (2012)
	main and secondary tasks	time sharing	Salthouse & Miles (2002)
	main and secondary tasks	time sharing	Hambrick et al. (2010)
	multiple tasks	switch between tasks	Hambrick et al. (2010)

The studies on multitasking usually defined multitasking with two different types of task strategies: sequential and parallel. For the sequential task strategy, multitasking is usually defined as the engagement of multiple task goals and the completion of individual tasks in sequence (Colom et al., 2010). The individual may, at any given point in time, be making progress towards meeting only one of the goals but over the longer time period makes progress towards all goals. This definition is based on the assumption that tasks are performed in a sequence, not simultaneously. Another definition of multitasking based on the sequential strategy is the engagement in individual and discrete tasks that are performed in succession (Dzubak, 2008). It is implied that there is necessary time spent switching between tasks. The task switching is an important part of the sequential processing of information and necessitates the selection of information that will be attended to, processed, encoded and stored.

Another type of task strategy is the parallel task strategy. In a parallel strategy, operator needs to handle all concurrent tasks at the same time (Bluedorn et al., 1992). In a study of the relationship between multitasking and academic performance, Junco and Cotton (2012) defined multitasking as divided attention and non-sequential task switching for the tasks in learning situations; for instance, a student is text messaging a friend while studying for an exam. This also emphasizes the ability of task switching between different activities. The ability of timesharing has been shown in many definitions of multitasking (Hambrick et al., 2010). Hambrick et al. (2010) defined multitasking as situations in which the performer must make conscious shifts of attention between two or more tasks (see also Spink et al., 2008), Oswald et al. (2007) described a model for investigating determinants of multitasking that orders potential predictor variables on a continuum from distal (indirect) to proximal (direct) causes.

For the purpose of this study, multitasking is defined as the engagement in multiple tasks that are performed simultaneously. Sequential operation of these tasks is not mandatory, but the ability to switch between different tasks is required for the procedure.

2.1.2 Multitasking and User Performance

Many aspects of user performance have been explored in the research on multitasking. This section introduces different types of user performance measures which have been investigated in multitasking environments.

Varying types of platforms have been provided in the research on multitasking performance. For instance, St. John, Harris, and Osga (1997) designed and compared two types of multitasking environments: multiple monitors and multiple windows to complete a flight simulation task. Kruchinin (2015) presented a multitasking system for pattern recognition. The system operates in real time, making it possible to achieve a maximum certainty value while solving several periodic tasks of pattern recognition. Speck-Planche and Cordeiro (2015) developed and examined multitasking models for quantitative structure-biological effect relationships to speed up drug discovery.

In media multitasking, David, Xu, Srivastava, and Kim (2013) examined communication multitasking in three conditions: instant messenger (IM) conversation with one partner, two IM conversations at the same time, and IM and phone conversation at the same time. Experiments were conducted with three conditions and results showed a higher task demand and a small loss in task performance of participants in the multitasking conditions. Perceived differences had not been noticed by multitasking participants who were focused on helping their partners. A strategic model of multitasking was suggested based on the results, with "IM being the preferred choice for tasks that require fewer, shorter exchanges and voice being the preferred choice for tasks that required more discussion and deliberation" (David et al., 2013, p.1658).

In the field of information behavior, multitasking information behavior has been compared with different personal variables. Psychologists have shown that information behavior is affected by: the affective domain, cognitive attributes, psychological factors, personality dimensions and sociological factors (Alexopoulou et al., 2014). Du and Spink (2011) presented a model to simulate multitasking, cognitive coordination and cognitive shifts on the web. However, this model does not incorporate personal variables and the impact of task or web design. Alexopoulou et al. (2014) addressed this gap in a study of the investigation of information behavior. This research explored the effects of working memory, cognitive coordination, cognitive shifts and various artifacts and task variables influenced by the PAT model (Personal, Artifact and Task characteristics) of flow. They recruited thirty university students and applied several assessment tools, including questionnaires, working memory tests, Flow State Scale test, think aloud data, observations, audio-visual data, web search logs and use of the Critical Decision Method. From the experiment, they evaluated the effects of these variables on multitasking and information behavior in the web. Also, an integrated framework was provided for a better understanding of information seeking behavior while using the web.

Flight and driving simulators are employed in a lot of research focused on multitasking performance. A popular system, the Multi-Attribute Task Battery (MATB), has been employed in many research studies as a multitasking platform in the past few decades (Comstock, & Arnegard, 1992; Miller, 2010; Miller, Schmidt, Estepp, Bowers, & Davis, 2014; Santiago-Espada, Myer, Latorella, & Comstock, 2011). An introduction to the history of this platform is provided in section 2.1.4. Here are some examples of the applications of MATB in previous studies.

In a study using a modified MATB, Singh, Sharma, and Singh (2005) examined the effect of training on workload in flight simulation task performance. A 2(training) x 2(session) x 3(block)

mixed factorial design was used. Results failed to show the effect of the amount of training on subjects' system-monitoring task performance under automated mode. Then Singh, Tiwari, and Singh (2009) examined the effects of automation training, automation reliability and workload on automation-induced complacency. 120 non-pilot participants operated a flight simulation task with several windows in which they had to detect automation malfunctions. The NASA-TLX was administered for the assessment of mental workload. A 2(training) x 3(reliability) x 2(session) x 3(block) mixed factorial design was used. Recorded measures of performances included hits, false alarms and reaction time on the system monitoring task and as root mean square errors on the tracking and on the fuel resource management tasks. The findings showed that perceived mental workload reduced from pre- to post-test sessions under high static reliability condition.

With the same multitasking platform, Singh, Tiwari, and Singh (2010) examined the effect of success and failure performance feedback on perceived mental workload and monitoring performance. They recruited 20 non-pilot subjects and recorded measures included hit rates, false alarms and root mean square errors of tracking task, and NASA-TLX rating scores. A 2(success-failure feedback) x 2(30-min sessions) x 3(10-min blocks) mixed factorial design was presented in this study. The results did not show a significant effect of performance feedback on mental workload and malfunction detection.

The ability of task switching has been observed and explored in the research on multitasking. Gutzwiller et al. (2014) developed a model of task overload management to evaluate task switching choices in their study. An experiment was conducted in MATB to explore the influence of two parameters of the model: task priority and task difficulty. They found that task difficulty and task priority predominantly and significantly influenced task switching.

Liu et al., (2016) examined the effects of WMC, task difficulty, and the ability of task switching in MATB-II (Santiago-Espada et al., 2011). A regression model was applied in this study to predict the relations among WMC, task switching ability, the level of task difficulty, and multitasking performance. The user performance measures they explored in their study included response time, response ratio, RMSD in tracking task, and subjective rating scores from NASA_TLX.

There are many other studies on multitasking that employed MATB. The performance measures covered a considerable range; for instance, reaction time, task error, efficiency, task speed, mental workload, and more. A summary of previous studies of performance measures in MATB is listed in Table 2.2.

Table 2.2 Summary of Research of Multitasking Performance in MATB

Author	Year	System	Task(s)	Independent Variables	Performance Measures
Phillips et al.	2007	MATB	SYSMON, COMM, TRACK	five levels of machine-initiated baud rate	human operator (HO) rate
Phillips et al.	2013	MATB	SYSMON, COMM, TRACK	five tasks, human operator (HO) rate	task-specific responses
Walters	2012	MATB	SYSMON, COMM, TRACK, RESMAN	Difficulty	weighting ratios, overall task complexity and human performance
Singh et al.	2010	modified MATB	Tracking	feedback (success and failure)	performance (hit rates, false alarms and RMSE), mental workload
Singh et al.	2009	modified MATB	Tracking	training (short and long), reliability	performance (correct detection, hit rate, false alarms, reaction time), mental workload
Singh et al.	2005	modified MATB	Tracking	training (short and long)	performance (correct detection, hit rate, false alarms, reaction time), mental workload
Bailey et al.	2006	modified MATB	SYSMON, COMM, TRACK, RESMAN	Condition (adaptive and yoke), System Reliability (high and low), Complacency Potential (high and low)	performance, gauge recall, mental workload, EEG

Table 2.2 Continued

Author	Year	System	Task(s)	Independent Variables	Performance Measures
Kidwell, Miller & Parasuraman	2014	AF_MATB	working memory task, four AF_MATB subtasks	task load, stimulation (active and sham)	Efficiency, tracking %time and RMSD
Durkee et al.	2013	AF_MATB	tracking	task difficulty	EEG, ECG, and eye-tracking activity, NASA TLX, performance measures
Bowers et al.	2014	AF_MATB	SYSMON, COMM, TRACK, RESMAN	task difficulty	EEG, ECG, EOG, NASA_TLX, and the shortened Dundee Stress State Questionnaire
Gutzwiller et al.	2014	MATB_II	SYSMON, COMM, TRACK, RESMAN	task priority, task difficulty	reaction time
Wilson, Lambert, & Russell	2000	MATB	SYSMON, COMM, TRACK, RESMAN	rest (base), low workload level, high workload level	Six EEG channels, electrocardiographic (ECG), electrooculographic (EOG).
Borghini et al.	2013 & 2015	Flight Simulator X, MATB	SYSMON, COMM, TRACK, RESMAN	task stage	EEG, HR, eye blinks rate (EBR)

Table 2.2 Continued

Author	Year	System	Task(s)	Independent Variables	Performance Measures
Onnasch, Ruff, & Manzey	2012	Microworld – MATB	SYSMON, COMM, TRACK, RESMAN	Reliability, block	Eye-tracking, Performance measures, Subjective measures
Smith & Gevins	2005	MATB	repetitive vigilance tasks, continuous working memory tasks, and MATB tasks	task difficulty	EEG, EOG, Mental Workload
Aricò et al.	2014	MATB	cruise flight phase, flight level maintaining, and emergencies	task difficulty	EEG, HR
Baldwin & Penaranda	2012	MATB	working memory tasks	task difficulty	classification accuracies

2.1.3 Physiological Measurements in Multitasking

"Physiological measures use the physical reactions of the body to objectively measure the amount of mental work a person is experiencing" (Miller, 2001, p.6). As compared with subjective ratings, the objective measures is much more exact method to assess mental workload from a person since no direct response is required from the person (De Waard, 1996). To assess the workload during multitasking, most research focused on five physiological areas: eye activity, cardiac activity, respiratory activity, speech measures, and brain activity.

Eye measures usually take horizontal eye movements, eye blink rate, and interval of closure into account for the assessment of mental workload (De Waard, 1996; Van Orden, Limbert, Makeig, & Jung, 1999). Speech measures mainly include pitch, rate, loudness, jitter, and shimmer (Brenner, Doherty, & Shipp, 1994). Cardiac activity is measured through heart rate (HR), heart rate variability (HRV), and blood pressure (Jorna, 1992; Roscoe, 1992 & 1993; Veltman & Gaillard, 1996 & 1998; Wilson, 1992). Respiratory activity measures the number of breaths in a given amount of time and the amount of air a person is breathing in (Roscoe, 1992; Wilson, 1992, 1993). To measure brain activity, either the electroencephalograph (EEG) or electrooculogram (EOG) are usually used (Fournier et al., 1999; Gevins & Smith, 2003).

Several workload measures use physiological changes in the eye to determine mental and visual workload (Miller, 2001). Although the eye is associated primarily with visual workload, it has been shown that some measures are able to accurately predict mental workload for some tasks as well (Van Orden et al., 2001). The main measures associated with the eye are horizontal eye activity (movement) (HEM), blink rate, and interval of closure. Other measures include eye fixation and pupil diameter. These eye measurements have been used in a variety of studies to

assess workload, especially during the operation in a multitasking environment. Table 2.3 shows a summary of some applications of eye measures in multitasking studies during the past decade.

Table 2.3 Summary of Eye Measures and Multitasking

Study	Eye Measurement	Primary Measure?	Other measures	System
Lin & Imamiya, 2006	Fixations number (FN), fixation duration (FD) and scanpath length (SPL).	No	NASA-TLX, heart rate variability (HRV), hand movements	NAC EMR-HM8, NAC Inc.
Imants & de Greef, 2014	Convex Hull Area, Fixation Clusters, Spatial Density, Nearest Neighbor Index	Yes	N/A	Eyelink II
Bailey & Iqbal, 2008	percent change in pupil size (PCPS), eye blink	Yes	N/A	Eyelink II
Ikuma, Harvey, Taylor, & Handal, 2014	Fixation Duration, time at AOIs (areas of interest)	No	Speed and Accuracy, NASA-TLX rating, Situatiuon Awareness Assessment (SAGAT)	EasyGaze
Di Stasi, Antolí, & Cañas, 2013	Saccade peak velocity, fixations, pupil diameter	No	Completion time, Working memory test	Eyelink II
De Rivecourt, Kuperus, Post, & Mulder, 2008	Saccade frequency, fixation duration/dwell time,	No	HR, HRV, rating scale of mental effort (RSME)	JazzManager
Hertzum & Holmegaard, 2013	Pupil Diameter	No	Percieved Time Ratio, NASA-TLX Ratings	SMI eye Tracker
Bhavsar, Srinivasan, & Srinivasan, 2015	Pupil diameter, fixation duration, saccadic duration, dwell time	Yes	response time	PCCR
Tokuda, Palmer, Merkle, & Chaparro, 2009	Saccadic Intrusions	Yes	N/A	Tobii 1750

Table 2.3 Continued

Study	Eye Measurement	Primary Measure?	Other measures	System
Dehais, Causse, & Pastor, 2008	Fixation Duration, Fixation Frequency, Pupil Diameter,	Yes	N/A	Tobii x50
Brouwer, Hogervorst et al., 2012	Pupil Size	No	RSME Ratings, Accuracy, EEG Recordings	Tobii T60 Eye Tracker Monitor, Labtec LCS-1050 Speakers,
Oksama & Hyönä, 2016	Fixation Frequency, Fixation Duration, AOI, Pupil size, Blink Rate, Saccades	Yes	N/A	Eyelink 2000
Benedetto, Pedrotti et al., 2011	Blink Rate, Blink Duration, Average Pupil Size	No	Subjective Ratings (NASA-TLX and RSME), Reaction Time	SMI iView X HED

A number of studies have examined the usefulness of the eye blink as an index of workload. Eye blink rate is the number of eye closures in a given amount of time and prior research has suggested that eye blink rate decreases as visual workload increases (Fogarty & Stern, 1989; Stern, Walrath, & Goldstein, 1984).

It was found that blinking may provide more information than just an estimate of workload (Stern et al., 1984). Stern and Skelly (1984) examined the effects of both visual and auditory information processing on blinking. It was found that blinking is "inhibited during the 'taking-in' of information, whether such information is presented visually or auditorily. Once a decision is made whether it requires action or requires the inhibition of action, a blink is likely to occur. The non-inhibition of blinking during the above mentioned time periods is associated with a higher likelihood of occurrence of missed signals and erroneous responses."

Fournier, Wilson, and Swain (1999) used the Multiple Attribute Task Battery to explore the mental workload through eye movement measures in their study. They had a single task (communication task) condition and three multiple task conditions that differed in workload. There was no resting baseline comparison, but the comparison of the single task to the multi-task conditions indicated that blink duration, rate and amplitude differentiated between these two conditions. Essentially, with multi-task workload the blink duration was shorter, the blink amplitude was greater and the blink rate was slower. However, a comparison among the three levels of multiple task workload found no differences. Performance measures discriminated among the multi-task workload level (as did cardiovascular indexes), so the lack of eye blink measures indicated poor diagnosticity and not a design failure to create distinct workloads.

Hankins and Wilson (1998) recorded eye blinks during a flight in a single engine aircraft. The flight was divided into 19 phases comprised of four basic categories: ground based preflight,

visual flight rules (VFR) and instrument flight rules (IFR) at high speeds. Blink rate varied over the phases of the flight, with blink rate lower during all of the IFR segments when the pilot wore goggles to block out visual input from outside of the cockpit. There was approximately a 50 percent decrease in blink rate under these conditions (10 blink/min difference). In comparison, the range of variation in the Verwey (1996) study was two blinks/min. Also, the flight segment requiring the pilot to perform a touch-and-go landing showed a lower blink rate, which was comparable to the IFR phases. This shows that eye blink is a moderately sensitive measure of workload. By comparison, a similar analysis of pilot heart rate more clearly differentiated among the flight segments and presumably different workload demands.

Backs, Ryan, and Wilson (1994) used a tracking task that factorially combined two levels of physical workload with three levels of perceptual/cognitive workload. Blink rate decreased from baseline levels during task performance. However, blink rate did not differ among the six tracking workload conditions. The authors also recorded respiration and heart activity measures which did differentiate among the workload conditions. This suggests that blink rate may be most useful as an index of the presence of workload, but not a good diagnostic choice for discriminating among levels of demand.

2.1.4 Multitasking Platforms

Many platforms have been employed in the study of multitasking and user performance (see Table 2.4). There are two general approaches used to develop the platform of multitasking in previous studies. One approach is to provide two or more systems and require the operator to perform various activities in these systems simultaneously. Another approach is to provide a system which will provide a set of multiple tasks for operator, for example, a system including both auditory and visual inputs. This study employed a computer-based task paradigm designed to evaluate operator performance and workload in multitasking, the Multi-Attribute Task Battery (MATB). The original MATB was developed by Raymond Comstock and Ruth Arnegard in 1992. MATB provides a benchmark set of tasks that are analogous to activities that aircraft crew-members perform in flight, but it can be operated by non-pilot subjects. The success of the MATB has led to several different versions, and this study used the MATB-II.

Air Force Multi-Attribute Task Battery (AF-MATB). In 2010, the AFRL's 711th Human Performance Wing's Human Effectiveness Directorate introduced an update to the original Multi-Attribute Task Battery (MATB) developed by Comstock and Arnegard (1992).

Second version of AF-MATB released in August, 2014. It added serial and digital port-triggering to AF-MATB, allowing the task to interface with neurophysiological acquisition systems. This change facilitated time-syncing between AF-MATB and acquired neurophysiological data and aided in the integration of state-based information into acquired data. Taking advantage of these changes requires sufficient knowledge of either electrical and/or biomedical engineering.

Table 2.4 Platforms of Multitasking

Single or Multiple System	Multitasking Platform	Tasks	Author & Year
Single	MATB	SYSMON, COMM, TRACK	Phillips (2007&2013)
Single	MATB	SYSMON, COMM, TRACK, RESMAN	St. John et al. (1997); Wilson (2000); Walters (2010&2012); Aricò et al. (2014); Borghini et al. (2015)
Single	MATB	working memory tasks	Baldwin & Penaranda (2012)
Single	modified MATB	flight simulation task	Singh et al. (2005, 2009, 2010)
Single	modified MATB	SYSMON, RESMAN, TRACK	Bailey et al. (2006)
Single	Microworld – MATB	SYSMON, COMM, TRACK, RESMAN	Onnasch et al. (2012)
Single	AF_MATB	tracking	Durkee et al. (2013); Bowers et al. (2014)
Single	MATB_II	SYSMON, COMM, TRACK, RESMAN	Gutzwiller et al. (2014)

Table 2.4 Continued

Single or Multiple System	Multitasking Platform	Tasks	Author & Year
Single	web	web searching	Alexopoulou et al. (2014)
Single	PRS	pattern recognition	Kruchinin (2015)
Single	Skype	instant message, skype	David et al. (2013)
Single	NASA BORIS simulation system	robotic arm control tasks	Li, Wickens, Sarter, & Sebok (2014); Wickens et al. (2015)
Single	UAV	low-fidelity UAV simulation task	Leidheiser & Pak (2014)
Single	simulated life-support system	simulated process control task	Clegg et al. (2014)
Multiple	AF_MATB, WMC	working memory task, four AF_MATB subtasks	Kidwell et al. (2014)
Multiple	MATB	simple repetitive vigilance tasks, continuous working memory tasks, and MATB (systems monitoring, resource management, communications, and tracking)	Smith & Gevins (2005)
Multiple	Flight Simulator X, MATB	SYSMON, COMM, TRACK, RESMAN	Borghini, Aricò, et al. (2013)
Multiple	Computer Based Assessment System	divided attention and funnel tasks	Colom et al. (2010)

*The Multi-Attribute Task Battery-II (*MATB-II) released in July, 2011. It added the collection of post-experiment workload ratings based on the NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988). MATB-II includes essentially the same tasks as the original MAT Battery, plus new configuration options including a graphical user interface for controlling modes of operation. MATB-II can be executed either in training or testing mode, as defined by the MATB-II configuration file. The configuration file also allows set up of the default timeouts for the tasks, the flow rates of the pumps and tank levels of the Resource Management (RESMAN) task. MATB-II comes with a default event file that an experimenter can modify and adapt.

As shown in Fig. 2.1, MATB-II consists of four main tasks (system monitoring, SYSMON; tracking, TRACK; communications, COMM; resource management, RESMAN) and two additional displays (scheduling, SCHED; pump status). In addition, a Workload Rating Scale (WRS) based on the NASA-TLX (Hart & Staveland, 1988) can be provided in a full window during the experiment.

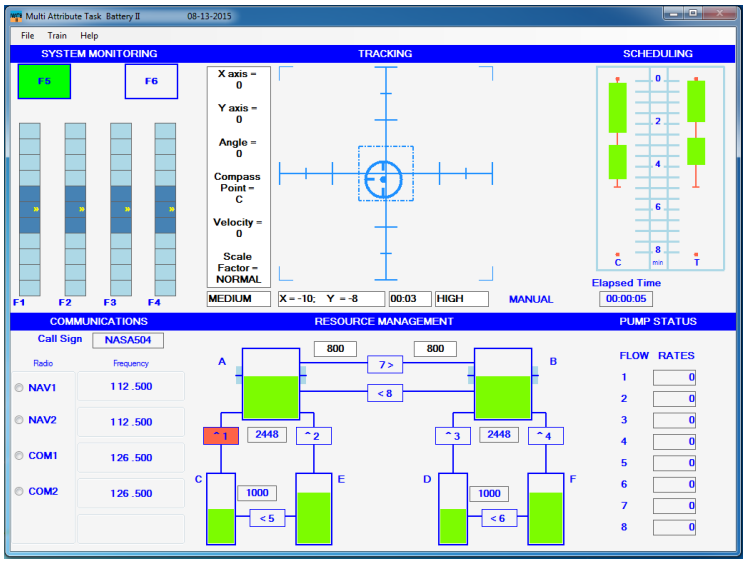
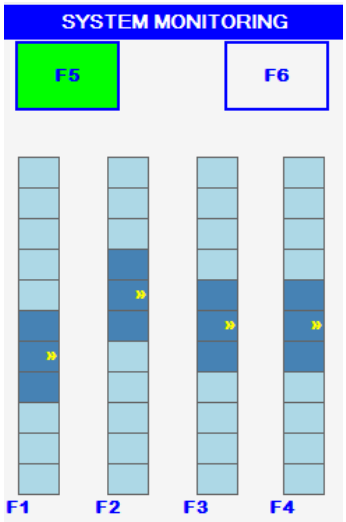


Figure 2.1 Multi-Attribute Task Battery II

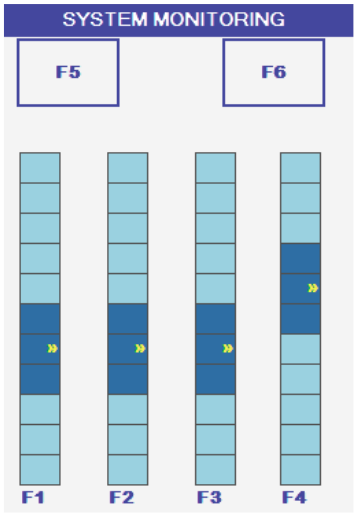
2.1.4.1.1 System monitoring (SYSMON) Task

The SYSMON task appears in the upper left corner of the MATB-II main window and is divided into two subtasks: warning lights and scales. There are two warning lights. The light on the left is normally green, indicating an “ON” state (see Figure 2.2 (a)). During the operation, if the light on the left turns “OFF” (from green to the background color), the subject should respond by clicking on it using the mouse or by pressing the “F5” key to turn it back “ON” (green) (see Figure 2.2 (b)). The light on the right is normally “OFF” (background color) (see Figure 2.2 (a)). When it turns “ON” (red), the subject is required to turn if “OFF” again by clicking on it using the mouse or pressing the “F6” key (see Figure 2.2 (c)).

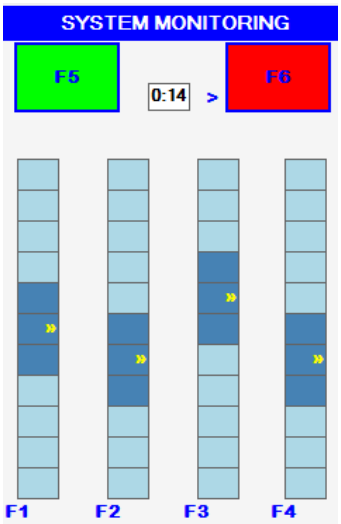
The second subtask of the SYSMON task includes four scales. There is an “indicator light” fluctuating around the middle of each scale. The subject is required to detect the scale whose indicator light shifts away from the middle of the scales and respond by clicking on the scale or pressing the function key indicated below the affected scale. Figure 2.2 (d) is an example of scale event: Scale Three moves up.



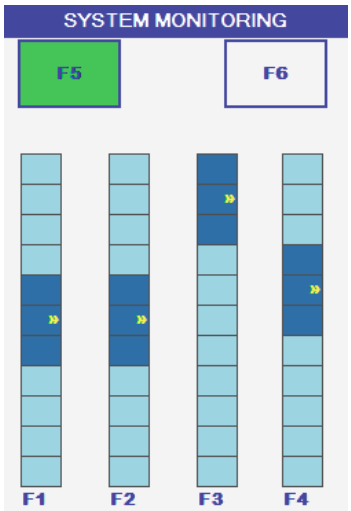
(a) SYSMON Normal State



(b) Green Light OFF



(c) Red Light ON

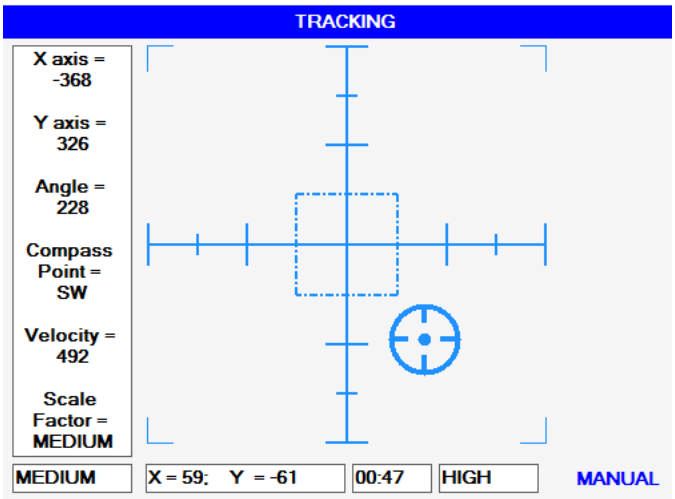


(d) Scale Three UP

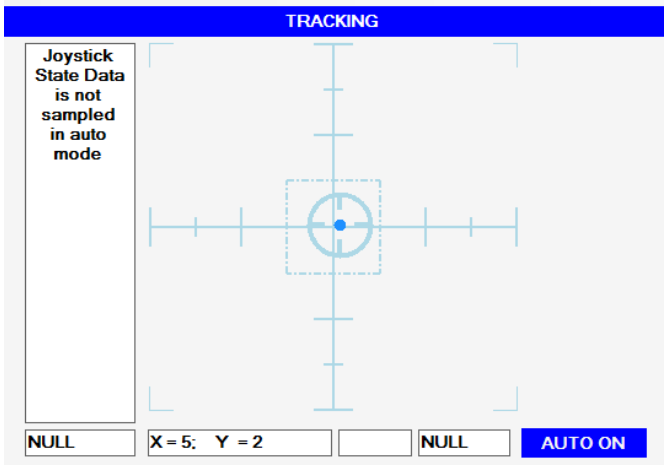
Figure 2.2 SYSMON Task

2.1.4.1.2 Tracking (TRACK) Task

The TRACK task is located in the upper center panel of the MATB-II window. This task has two modes: manual and automatic. In the manual mode, the subject is required to use a joystick to keep the circular cursor within the center of the inner box (see Figure 2.3 (a)). In the automatic mode, the circular cursor is controlled by MATB-II and will fluctuate and remain in the inner box (see Figure 2.3 (b)).



(a) Manual mode

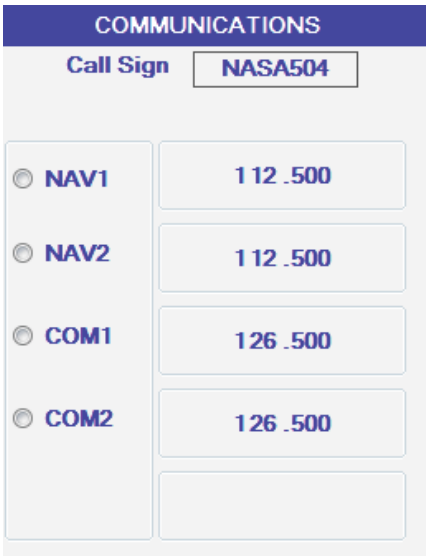


(b) Automatic mode

Figure 2.3 TRACK Task: (a) Manual and (b) Automatic mode.

2.1.4.1.3 Communications (COMM) Task

The COMM task is located in the lower left corner of the MATB-II main window. Figure 2.4 shows the initial state of COMM task. During the operation, one or more audio messages with particular "Call Sign" will be played which require the subject to change the frequency (right side) of a specific radio channel (left side). Figure 2.4 (b) shows an example of COMM task: radio channel COM1 event is handled.



(a) COMM Initial State

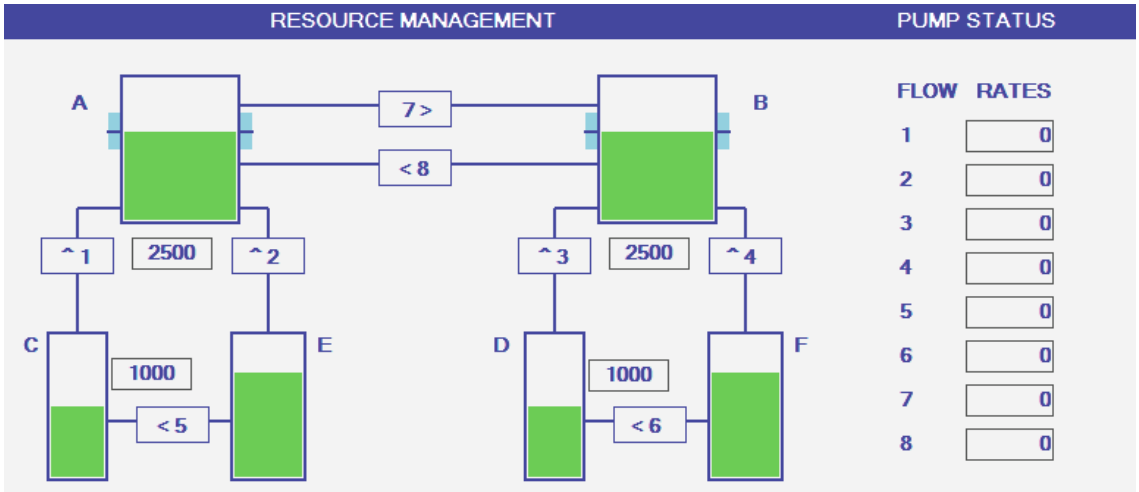
(b) COMM COM1 Event Handled

Figure 2.4 COMM Task

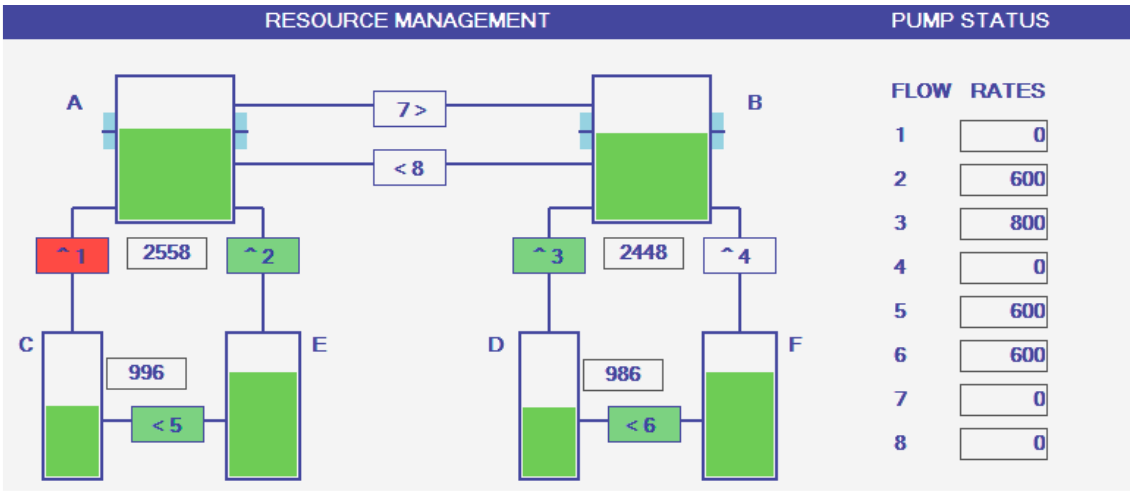
2.1.4.1.4 Resource Management (RESMAN) Task and Pump Status

The RESMAN task is located in the bottom center of the MATB-II main window. In this window, there are six tanks labeled with letters A – F and eight pumps labeled with numbers 1 – 8. During this task, the subject should maintain the level of fuel in tanks A and B within ±500 units of the initial condition of 2500 units each by manipulating the pumps (see Figure 2.5 (a)). The initial state of each pump is colored in gray. If one pump is selected and activated, its color will become red (see Figure 2.5 (b)).

On the left-hand side, there is a "Pump Status" window shows the flow rates for all eight pumps in RESMAN task (see Figure 2.5). This section does not require the operation from subject.



(a) RESMAN Initial State



(b) RESMAN Activated

Figure 2.5 RESMAN Task and Pump Status

2.1.4.1.5 Scheduling Display (SCHED)

The Scheduling display (Figure 2.6) is located in the upper right panel of the MATB-II main window. SCHED allows the subject to "look ahead" for up to eight minutes in the future of COMM ("C" at left side) and TRACK ("T" at right side) tasks. At the bottom of this display, an elapsed timer tracks the time spent of a MATB-II trial. The SCHED display does not require intervention from subject. It only provides "look ahead" information for COMM and TRACK tasks and there is no data recording from this display.

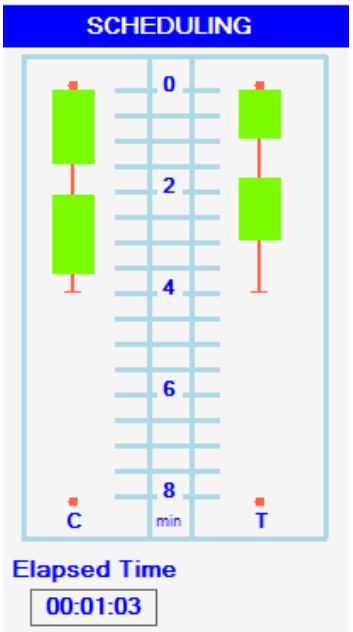


Figure 2.6 SCHED Display

Note: C represents COMM task; T represents TRACK task; Elapsed Time shows the elapsed time since the start of the trial.

2.1.4.1.6 Workload Rating Scale (WRS)

If specified by the script file, a Workload Rating Scale will appear in a full window during the operation of MATB-II (Figure 2.7). This WRS is based on the NASA Task Load Index (NASA-TLX) (Hart & Staveland, 1988). It includes six subscales: mental demand (MD), physical demand (PD), temporal demand (TD), and the individual's perceived level of performance (PE), effort (EF), and frustration (FR). The WRS intends to estimate subject's overall mental workload during the operation. An instruction of the WRS is included in Appendix C.

When the WRS is activated, the run elapsed timer will be paused until the subject saves the ratings or the predefined presentation time for WRS is reached. Once the WRS session ends, the run elapsed timer resumes.

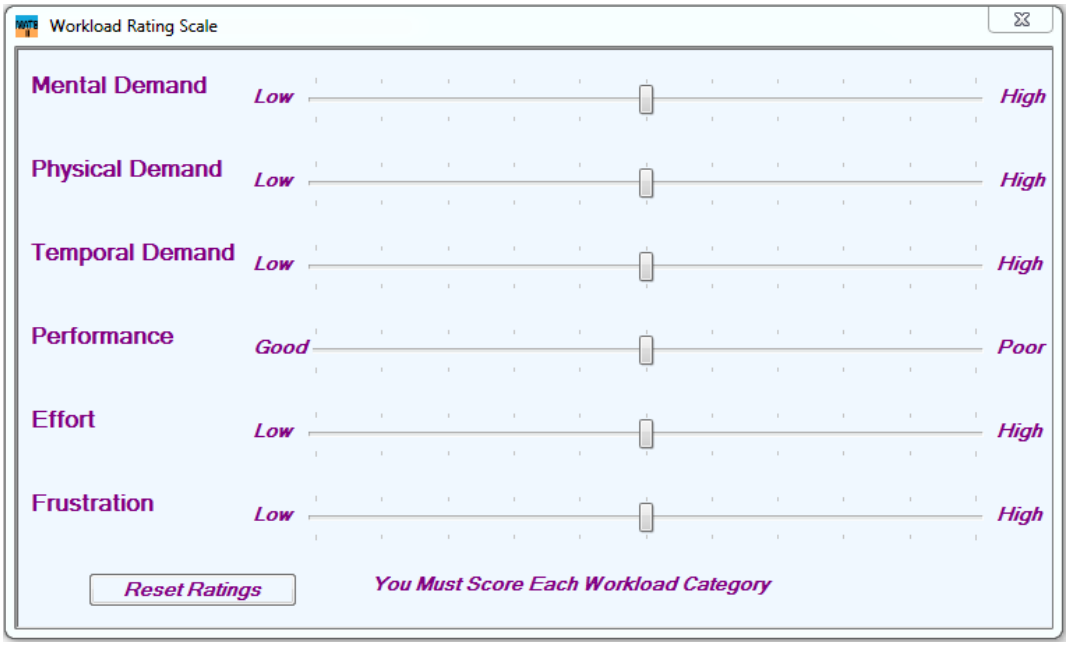


Figure 2.7 The Workload Rating Scale

2.2 User Performance Modeling

2.2.1 Quantification of Information

In the area of human information processing in multitasking, many studies described the human operator as a transmitter of information in (Wickens, 1991, 2008; Wickens et al., 2015). Formally, information is defined as the reduction of uncertainty (Shannon & Weaver, 1949). This introduction of information theory (Shannon & Weaver, 1949) allowed psychologists in the early 1950's to explore human performance in this framework. The classical works based on information theory include Hick (1952), Hyman (1953), and Fitts' law (1954). Beyond these early contributions in human information processing, researches on information theory had been extended into many directions and situations, for example, the assessment of information channel and user performance in multitasking environments (Camden et al., 2016; Chan & Childress, 1990; Phillips et al., 2007; Repperger et al., 1995). The main contribution of these implications of information theory is the quantification of information transmitted from the system as a baud rate (bits per second) (Crossman, 1960; Repperger et al., 1995).

To have a better understanding of the models of quantifying information, this section provides a review of basic concepts of the information theory (Shannon, 1948) and Fitts' law (1954), as well as their applications on human information processing.

2.2.1.1 History and Applications of Information Theory

In 1948, Claude E. Shannon published a mathematical theory of communication, which laid the foundation of information theory. He proposed a fundamental model for communication systems which included five essential parts: (1) an information source which produces a message or sequence of messages to be communicated to the receiving terminal; (2) a transmitter which

operates on the message in some way to produce a signal suitable for transmission over the channel; (3) the channel is merely the medium used to transmit the signal from transmitter to receiver; (4) the receiver ordinarily performs the inverse operation of that done by the transmitter, reconstructing the message from the signal; (5) the destination is the person (or thing) for whom the message is intended. Figure 2.8 shows the original schematic diagram of a general communication system that Shannon (1948) proposed.

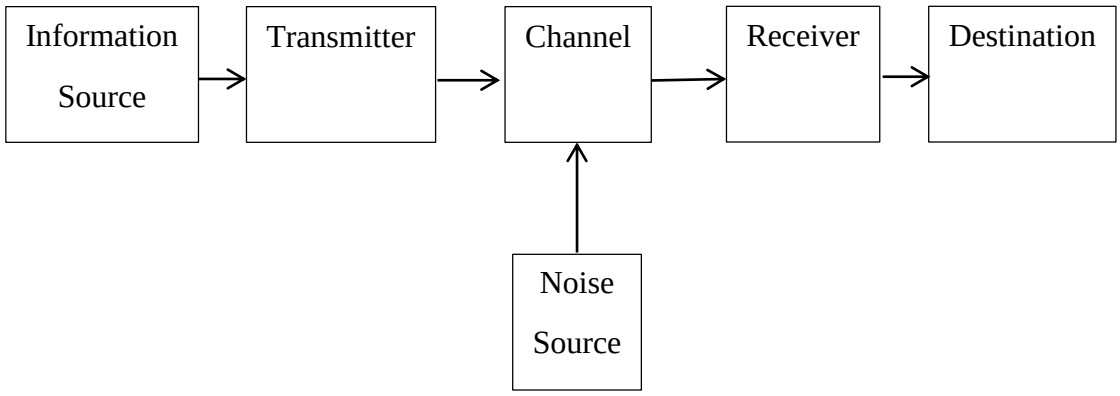


Figure 2.8 Diagram of a general communication system (Shannon, 1948)

The information source produces a message or a sequence of messages that are communicated to the receiver. The transmitter operates on the messages produced by the information source and in the same way produces a signal suitable to be transmitted. Comparing information theory to a MATB-II system, the receiver will be replaced by a human operator and the destination is the permanent record of the human information output.

There are three categories of a general communication system (Shannon, 1948): discrete, continuous and mixed systems. Discrete systems are produced when the message and signal are a

sequence of discrete symbols. Continuous systems have the message and signal as a stream of events. Mixed systems contain both discrete and continuous variables. MATB-II is a mixed systems.

Limitations of Shannon's information theory have been discussed in some studies. Shannon's definition of information is not absolute and the probability in the expression of information may be a random variable in practice, but in information theory the probability is treated as a fixed value (Wang, 2009). Shannon provided the definition to entropy of the joint events and conditional entropy, produced the formula, and made the conclusion that the uncertainty of y would be never increased by knowledge of x and it would be decreased unless x and y were independent events, in which case it was not changed. Wang (2009) argued that Shannon's conditional entropy is not suitable to be called a conditional entropy since the uncertainty of y is never increased by knowledge of x, and is not absolute.

2.2.1.2 Applications of Fitts' law

Fitts' law is a model of human psychomotor behavior derived from Shannon's Theory, and it is a fundamental theorem of communication systems (Fitts, 1954; Shannon & Weaver, 1949). The realization of movement in Fitts' model is analogous to the transmission of information. Fitts (1954) suggested that the difficulty of a task could be measured using information metric bits. Additionally, he introduced the idea that, in carrying out a movement task, information is transmitted through a channel - a human channel, and is said to transmit in bits. If the number of bits is divided by the time to move, then a rate of transmission in "bits per second" can be ascertained (MacKenzie, 1992).

This model established the information capacity of the human motor system. This capacity, which is called the index of performance (IP) in the model, is analogous to channel capacity (C)

in Shannon's theory. IP is calculated by dividing a motor task's index of difficulty (ID) by the movement time (MT) to complete a motor task. Thus,

$$IP = ID/MT$$

Fitts claimed that electronic signals are analogous to movement distances or amplitudes (A) and that noise is analogous to the tolerance or width (W) of the region within which a move terminates. Then, the index of difficulty for a motor task was expressed as:

$$ID = \log_2\left(\frac{2A}{W}\right)$$

Building on Fitts' evidence that the rate of human information processing is constant across a range of task difficulties, a variety of research has examined the use of Fitts' Law in both computerized and real environments for different limb and muscle groups (MacKenzie, 1992; Mateo, 2005; Murata & Iwase, 2001). Different input devices, such as mouse, trackball, joystick, touchpad, helmet-mounted sight, and eye tracker, have been tested by researchers. Fitts' law has performed well with the velocity-control and position-control isometric joysticks in many studies (Bharwani, 2006; Epps, 1986; Jagacinski & Monk, 1985; Card, English, & Burr, 1978; Kantowitz & Elvers, 1988).

2.2.2 Human Information Processing Models

Extensive studies have been done in the domain of human information processing modeling and different types of quantitative models have been developed in the studies of human information processing. This section reviewed several quantitative models for user information processing.

To get a comprehensive understanding of human information processing, Kieras and Meyer (1995) constructed several models for human multitasking performance by following the EPIC (Executive Process-Interactive Control) architecture. In their study of a classic dual-task with tracking and decision making tasks, two models were formulated based different task strategies, a lockout model and a interleave model. In the lockout model, tasks were honored with different levels of priority and the task with a lower priority was "locked out" (suspended) from the other task with higher priority. There was no overlap allowed of the executive process on the two tasks. On the contrary, the strategy of the interleave model was to assign equally priority to the two tasks and overlaps between these two tasks were allowed as much as possible though task operation. From the results, they revealed that the interleave model fitted more accurate to the empirical data to predict user performance in dual-task comparing to the lockout model.

By using the EPIC cognitive architecture, Zhang and Hornof (2014) constructed individualized computational cognitive models to demonstrate complexity of multitasking behavior in their study. They proposed an approach to investigate various "microstrategies" and applied this approach in a visually interleave dual-task paradigm. Individualized parameters were selected for a highly-interactive dual task including tracking and choice selection, such as text recoding time, speed and direction recoding time, tracking coefficients, glance interval, etc. Under the time-pressured multitasking environment, they suggested that multitasking performance will be affected not only by parallel cognitive and sensorimotor processing, but also by the selection of subtask microstrategies.

GOMS (Goals, Operators, Methods, and Selection rules) models were first proposed by Card, Moran, and Newell (1983) to describe how operators carry out tasks within a certain system. Cognitive task analysis (CTA) intended to capture a description of the explicit and implicit

knowledge that experts use to perform complex tasks through a variety of interview and observation strategies (Clark & Estes, 1996). It is defined as "interview and observation protocols for extracting implicit and explicit knowledge from experts for use in instruction and expert systems" (p.578).

In the field of robotics and automation, a cognitively plausible GOMSL (Goals, Operators, Methods, Selection rules) model was demonstrated by Kaber et al. (2006). In their study, human performance in robotic rover control was represented by GOMSL model and comparisons were made with actual human performance. The results showed that GOMSL code produced more precise control of the rover than human performance, but this was at the cost of time. Based on the initial model (Kaber et al., 2006), Kaber, Kim and Wang (2011) revised the GOMSL model for a better understanding of human behavior and strategy in robotic rover control. Meanwhile, Kaber et al. (2011) also constructed computational GOMSL (Goals, Operators, Methods, and Selection rules Language) models based on a combination of cognitive task analyses (abstraction hierarchy modeling and goal-directed task analysis). Similar conclusions of the GOMS models were addressed in these studies: the GOMS modeling approaches "have potential for describing interactive closed-loop rover control with continuous monitoring of feedback and corresponding control actions" (Kaber et al., 2011, p.53).

2.2.3 Task Difficulty

The manipulation of task difficulty has been applied in many studies as suggested that when an increase in task difficulty occurs in a system, more resources from the human operator will be required to maintain performance at an optimal level (Wickens, 1991). According to performance-resource function (Norman and Bobrow, 1975), the effect of task difficulty is easy to describe and reasonably straightforward. During dual-task performance, with the increase of task difficulty, more resources will be required from the human operator to maintain task performance at an acceptable level.

To estimate user performance in different multitasking conditions, task difficulty is an efficient parameter to predict workload during experiment which can be predefined by designer (Adler & Benbunan-Fich, 2015). To examine how different types of multitasking and different levels of subjective task difficulty influence performance, Adler & Benbunan-Fich (2015) conducted a controlled experiment using a custom-developed multitasking environment with a total of 636 participants. Three conditions of multitasking were designed in the experiment, including discretionary, mandatory, and sequential. One primary task and five secondary tasks were completed by each participant. Subjective task difficulty was assessed by a 5 points rating scales. Significant effects of subjective task difficulty were found in user performance of both primary and secondary tasks.

For the input device of a system, the usability can be evaluated as a function of task difficulty. Rupp, Oppold, and McConnell (2014) conducted two experiments to compare the preference of game controllers in a multitasking environment, MATB-II, with the manipulation of task difficulty during the experiments. Two levels of task difficulty of the tracking task were manipulated in the experiments to control mental workload, but the details of the manipulation of

task difficulty were not provided. The results showed that game controllers are highly usable input devices and do not require high mental workload to operate, and they concluded that the game controllers are suitable for complex control tasks.

Task difficulty in single-task paradigms has been evaluated in the studies on brain activity and human information processing (Boiten, Sergeant, & Geuze, 1992; Van Winsun, Sergeant, & Geuze, 1984) and increasing studies have applied the manipulation of task difficulty in multitasking environments to investigate the brain activity and cognitive processing (Fournier et al., 1999; Gundel and Wilson, 1992). Fournier et al. (1999) examined whether alpha ERD and theta ERS could successfully measure workload of a complex, single event while a subject was engaged in multiple, interactive tasks. To achieve this goal, they manipulated task difficulty in a multitasking system, MATB, and recorded the subjective rating of task difficulty during the experiments to validate the measures related to workload. Three levels of task difficulty were manipulated by changing number of stimuli in MATB tasks. They found that alpha 2 EEG, heart rate, behavioral, and subjective measures were sensitive to changes in workload in multitasking. They also revealed effective indexes of cognitive and behavior demand, such as Alpha ERD, theta ERS, eye blink, and behavioral measures.

Task difficulty has been evaluated with other predictors of multitasking performance in previous studies. Liu et al. (2016) investigated the effects on task difficulty, task switching, and WMC in a multitasking environment, MATB-II. They manipulated three levels of task difficulty in the MATB-II tasks with participants have different levels of task switching ability and WMC. A regression model was applied to analyze the effects of task difficulty and other two predictors on MATB-II measures. Their results indicated significant effects of all three predictors in multitasking performance.

2.3 Individual Differences and Practice in Multitasking

2.3.1 Effects of Working Memory Capacity on Multitasking Performance

Working memory refers to the temporary storage and manipulation of the information necessary for a wide range of complex cognitive activities (Baddeley & Hitch, 1974). Baddeley and Hitch (1974) proposed a three-component working memory model and emphasized the dynamic interaction of memory maintenance and attention control in complex cognitive tasks. It has been assumed to be an important feature in multitasking performance (König et al., 2005). Working memory capacity (WMC) has been shown to predict performance in various cognitive tasks, e.g., following directions, writing, note-taking, computer language learning, and bridge playing (Engle & Kane, 2004; Redick & Engle, 2006; Unsworth et al., 2005).

In the studies of cognitive abilities, individual differences in WMC has been evaluated through different types of span tasks, such as counting span, operation span, and reading span tasks (Conway, Kane, Bunting et al., 2005). The operation span (OSPAN) task is originally developed by Turner and Engle (1989) and widely used measure of WMC (Conway et al., 2005; Unsworth et al., 2005). The OSPAN task requires subjects to remember words or letters while solving mathematical operations (Turner & Engle, 1989). Different versions of the OSPAN task have been developed during the past decades (e.g., Engle, Cantor, & Carullo, 1992; Unsworth et al., 2005) to help the researches on screening participants with different WMC span groups. There are limited studies that applied OSPAN task as a screen method and published their criteria for WMC screening. Table 2.5 summarized these studies and listed the cut-off criteria of high- and low-span groups employed by the previous studies.

Table 2.5 Cut-off Criteria in OSPAN Tasks

Study	Score range of OSPAN	Cut-off scores (high/low)	Population pool	Number of subjects (high/low)
Kane et al., 2001	0-60	N/A	large pool of undergraduates	107/96
Kane & Engle, 2003	0-60	19/9	300-400 undergraduates	40/47
Goldinger et al., 2003	0-54	21/12	138 undergraduates	35/35
Watson et al., 2005	N/A	20/9	undergraduates	50/50

Kane and Engle (2003) investigated individual differences in WMC and performance on the Stroop task in their experiments and explored that the attentional, "executive" component of the working memory model (Baddeley & Hitch, 1974) was specifically responsible for the covariation between measures of working memory and higher order cognition. They first screened participants for WMC with the OSPAN task (La Pointe & Engle, 1990) and tested participants from two groups, high spans (the top quartile of the distribution in OSPAN task) and low spans (the bottom quartile of the distribution in OSPAN task), with Stroop task. They discovered that the proportion of congruent trials in the Stroop task drove working memory span differences in error interference. For example, when the Stroop task included large number of congruent trials (red presented in red), individuals from the low span group committed more errors than did individuals from high span group on the rare incongruent trials (blue presented in red). They found that WMC was related to latency, not accuracy in the Stroop task. They suggested that Stroop

interference is jointly determined by two mechanisms, goal maintenance and competition resolution, and both of them depend on WMC.

Based on cognitive psychology research, many predictors of multitasking performance have been discussed in previous studies. König et al. (2005) explored five different predictors of multitasking performance in a computerized scenario which consisted of number, letter, figure, and question components (called "Simultaneous capacity/Multi-tasking" or SIMKAP). The predictors that they tried to test included working memory, attention, fluid intelligence, polychronicity, and extraversion. To guarantee durable results, they chose three tests to measure working memory: Reading Span, Spatial Coordination, and Switching Numerical (Oberauer, Süß, Schulze et al., 2000). The Reading Span task and Spatial Coordination can assess the working memory component "storage of information in the context of processing" with verbal material and numerical material, respectively. The Switching Numerical test can measure the supervision component of working memory with numerical material. Results of 122 participants showed that working memory was the most important predictor in addition to attention and fluid intelligence. Even though attention was the most basic construct they used as a predictor, working memory showed the highest correlations with the multitasking measures in SIMKAP and explained incremental variance that could not be accounted for by other predictors.

In a study conducted with Multi-Attribute Task Battery II (MATB-II) (Liu et al., 2016), individual differences in WMC was examined with other multitasking predictors (e.g., the ability of task switching, task difficulty). Three sessions were completed by each participant: an automated OSPAN task to measure their WMC (Unsworth et al., 2005), a set of Trail Making Tasks (TMT) to estimate their task switching ability (Sanchez-Cubillo et al., 2009), and a set of Multi-Attribute Task Battery II (MATB-II) tasks to investigate their performance in multitasking

(Santiago-Espada et al., 2011). A regression model was applied in this study to predict the relations among WMC, task switching ability, the level of task difficulty, and multitasking performance. The results indicate that these three predictors had significant effects on multitasking performance and there was a significant correlation between these predictors and user performance. Users with a higher level of WMC (OSPAN scores) showed a better overall performance in MATB-II and experienced less mental workload during the experiment.

Some jobs might require both quick and correct multitasking performance, and some jobs might be especially important not to make errors (e.g., pilots) (Bühner, König, Pick, & Krumm, 2006). To explain the role of working memory in predicting the speed and the error aspect of multitasking, Bühner et al. (2006) conducted another experiment using SIMKAP (König et al., 2005) based on a multidimensional model of working memory including storage in the context of processing, coordination, and supervision (Oberauer, Süß, Wilhelm, &Wittmann, 2003). Attention and reasoning were controlled when they predicted multitasking speed and error. Each of the 135 participants completed a working memory test (Oberauer et al., 2003), a reasoning test (Amthauer, Brocke, Liepmann, & Beauducel, 2001), 2 attention tests (Zimmermann & Fimm, 2002), and a multitasking scenario (SIMKAP). As expected, working memory was the best predictor of multitasking performance in their study. Also, working memory components showed a differential validity when predicting speed and error in multitasking: multitasking speed was predicted mainly by coordination, and multitasking error was predicted mainly by storage in the context of processing. The results proved that reasoning, attention, and the working memory component supervision were less important or even unimportant predictors of multitasking speed and error.

2.3.2 Practice and Multitasking Performance

Practice and education are designed to improve learning and produce qualified performance which is required in many professions. As cognitive workload increases with the increase in task complexity or the number of tasks required to be performed at once, task performance deteriorates sharply. Fortunately, previous research showed that extensive training and practice can greatly reduce such deterioration (Dux et al., 2009). Improvements in task performance are generally achieved by qualitative changes in cognitive processes. An assumption held frequently in cognition research on skill acquisition is that performance improvements due to practice are best fit by negatively accelerated power functions, called power law of practice (Johnso, Bellman, & Lohse, 2003; Newell & Rosenbloom, 1981; Palmen, 1999; Palmeri, 1997) and the power law of practice is often used as a benchmark test for theories of cognitive skill acquisition (Haider & Frensch, 2002; Palmen, 1999; Palmeri, 1997). Several studies investigated the effect of practice on the performance in single or dual tasks.

Basak, Boot, Voss, & Kramer (2008) applied training of video games and showed that cognitive training can improve user performance and attenuate cognitive decline in older adults. They mentioned that declines in various cognitive abilities, particularly executive control functions, were observed in older adults and one important goal of cognitive training is to slow or reverse these age-related declines. They recruited 40 older adults and evenly separated them as a training group and a control group. A battery of cognitive tasks were assessed before, during, and after video game training, including tasks of executive control (Operation Span, Task Switching, N-back task, Visual Short-Term Memory, Ravens Advanced Matrices, and Stopping Task) and visuospatial attentional tasks (Functional Field of View, Attentional Blink, Enumeration, and Mental Rotation). Results of the training group showed a significant improvement in the measures

of game performance and in a subset of the cognitive tasks, such as task switching, working memory, visual short term memory, and mental rotation than the control group. However, it is not known that whether video game training improves performance on everyday cognitive abilities and real-world tasks, such as driving skills, working in a busy office, or pilot skills.

Practice can lead to the increase in performance efficiency and has been explored with other features in multitasking. Murray and Bellman (2011) explained the effect of practice, prior knowledge and task difficulty on consumer demand for hedonic experiences. Two groups, 207 and 114 participants, had performed an online video game with two levels of task difficulty (slow game vs. fast game) and two levels of the practiced number (1 game vs. 10 games). The results of practice showed a significant positive correlation with demand and indicated that practice significantly increased game playing efficiency. This study provided an evidence that practice improves user performance within a given period of time over repeated experiences.

Another study focused on the effects of practice and other task features such as connected skill learning, practice, and measures of performance (Stafford & Dewar, 2014) also recruited video game players. They had a very large sample size (N = 854,064) to explore the relationships. The online game included rapid perception, decision making, and motor responding. Stafford and Dewar (2014) extracted the game data and found that the lawful relations existed between practice amount and subsequent performance, and between practice spacing and subsequent performance. As players practiced, their average score improved.

The interesting issue of whether the training skill is transferable to other domain skills was not mentioned and addressed in most studies. Sutter, Oehl, and Armbrüster (2011) investigated the skill transformations and evaluated practice and carryover effects on the use of small interaction devices in an applied text-editing task. The touchpad and mini-joystick were used as small motion-

and force-controlled interaction devices in their study. They evaluated the performance differences between experts and novices. They found that the efficient performance of experts depended on domain-specific skills and they were not transferable. In addition, novices could achieve higher levels of performance with considerable. And they concluded that training procedures is essential for complex applications and/or unfamiliar device transformations.

Training-induced changes in brain activity have been examined (Dux et al., 2009; Erickson et al., 2007; Maclin, Mathewson, Low et al., 2011). While it is not known how training alters the brain to solve the multitasking problem, Dux et al. (2009) showed that training reduced the reaction times to each task under both single- and dual-task conditions in a behavioral training in their study. They examined several neural mechanisms that might have influence on training in multitasking. The findings showed that training will increase the speed of information processing in the related brain region which will reduce multitasking interference, then, multiple tasks will be processed in rapid succession. The findings revealed how training leads to efficient multitasking and provides evidence of multitasking limitations, called the poor speed of information processing in human prefrontal cortex.

3 Research Plan and Hypotheses

In the previous chapter, a series of relevant research was reviewed. Among a variety of cognitive models developed to assess user performance in multitasking environments, only a limited number of studies have quantified the information content transmitted between systems and human operator and even fewer studies have quantitatively evaluated individual differences in their models. The limitations of previous studies provided significant direction for the development of the present study.

A quantitative approach was proposed to investigate multitasking performance in this research. Chapter 4 will introduce the proposed quantitative model. The proposed model included quantification of stimuli as baud rate (bits per second), selection of task difficulty and task weight, and rearrangement of task weights.

Then, two phases of experimentation were conducted. The objectives of the first phase were to apply the proposed model in a multitasking environment and identify a performance baseline for each participant. The second phase aimed to validate the proposed model by rearranging a set of multitasks and comparing user performance with the baseline performance obtained in the first phase. The effects of individual differences in WMC and task difficulty were investigated at the end of the experiments.

3.1 Phase I

During the first phase, a baseline performance for each participant was obtained in a multitasking environment, MATB-II. Figure 3.1 shows the procedure of Phase I experiment.

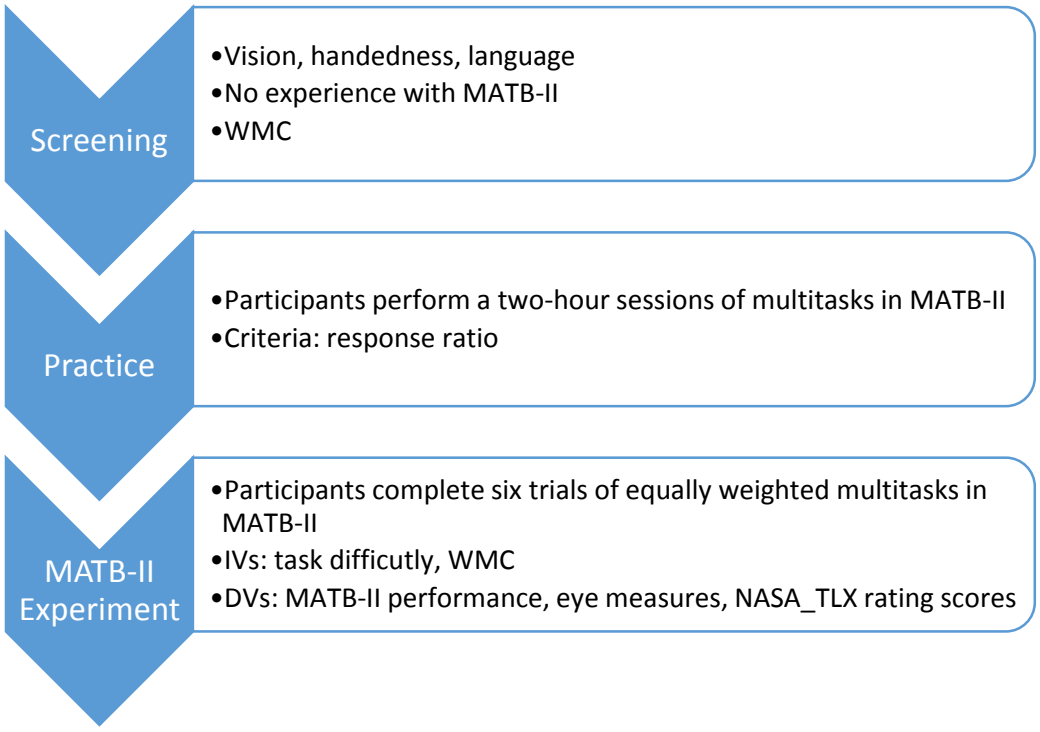


Figure 3.1 Experiment Design in Phase I

At the initial contact, an online screening session was performed by the potential participants to verify their age, native language, health condition, and handedness, as well as to estimate their WMC. During the practice session, participants were required to complete a two-hour training. If their multitasking performance met the predefined criteria, they were selected to undertake another MATB-II experiment session and the Phase II experiment. If they did not meet

the performance criteria, they were not selected to participate following experimental sessions. During the experiment, MATB-II performance (response time (RT), response ratio (RR), root-mean-square deviation (RMSD)) and mental workload (NASA-TLX rating scores and eye movement measures) were recorded. A baseline performance for each participant was calculated at the end of the first phase.

3.2 Phase II

During the second phase, a rearranged set of multitasks was completed by each participant and their performance was compared with the baseline performance from the first phase to validate the proposed model. Figure 3.2 shows the procedure of Phase II experiment.

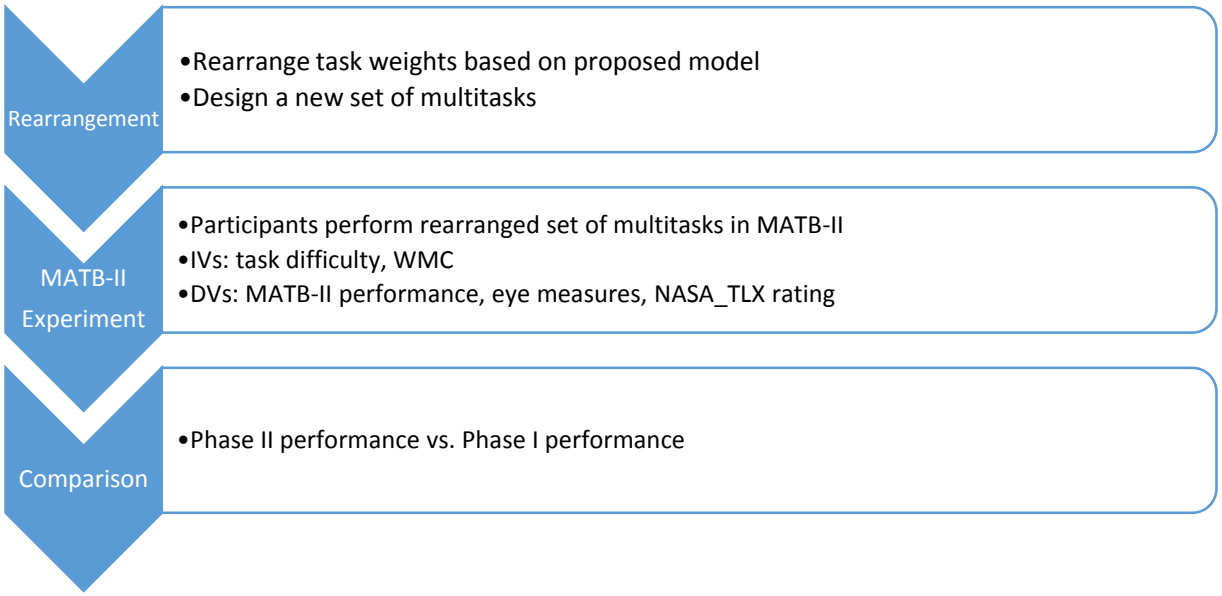


Figure 3.2 Experiment Design in Phase II

Based on the baseline performance from each participant and the proposed model, a new set of multitasks with different task weights was designed for each participant, respectively. The same participants who completed the Phase I experiment were invited to complete the Phase II experiment. During the experimental session, MATB-II performance (RT, RR, and RMSD) and mental workload (NASA-TLX rating scores and eye movement measures) were recorded. Phase II performance was compared with Phase I performance at the end of the experiments.

3.3 Hypotheses

There were three sets of hypotheses formulated for this study:

H1: User multitasking performance will improve after the rearrangement of multitasks.

H1 (a): The response time of MATB-II tasks is lower in Phase II than in Phase I.

H1 (b): The RMSD of TRACK tasks is lower in Phase II than in Phase I.

H1 (c): The mental workload of operators is lower in Phase II than in Phase I.

H2: Task difficulty will affect user performance in a multitasking environment.

H2 (a): Higher task difficulty will decrease the overall response ratio in MATB-II tasks, in both phases.

H2 (b): Higher task difficulty will increase the average response time in MATB-II tasks, in both phases.

H2 (c): Higher task difficulty will increase the RMSD in the MATB-II tracking task, in both phases.

H2 (d): Higher task difficulty will increase the mental workload of operators, in both phases.

H3: Participants with different working memory capacities (WMC) will have different performances in a multitasking environment.

H3 (a): Participants with higher WMC will have a higher overall response ratio in MATB-II tasks than participants with lower WMC.

H3 (b): Participants with higher WMC will have a lower average response time in MATB-II tasks than participants with lower WMC.

H3 (c): Participants with higher WMC will have a lower RMSD in the MATB-II tracking task than participants with lower WMC.

H3 (d): Participants with higher WMC will have a lower mental workload in MATB-II tasks than participants with lower WMC.

4 The Proposed Quantitative Model for User Performance in Multitasking

4.1 Baud Rate Selection and Task Difficulty

To quantify the stimuli generated from a multitasking system, this model converted the information content of the stimuli into a baud rate (bits per second or bps) based on Shannon's information theory (1948) and previous studies (Camden et al., 2016; Liu et al., 2016; Phillips et al., 2007; Repperger et al., 1995). Baud rate was defined as "spatial information, in bits, divided by temporal information, in seconds" (Camden et al., 2016, p.137) in this study, representing the rate that information was transmitted in a communication channel, measured in bps.

Bit is "a unit of computer information equivalent to the result of a choice between two alternatives" (Merriam-Webster, retrieved 2017). In information theory, one bit is typically defined as "the uncertainty of a binary random variable that is 0 or 1 with equal probability", or "the information that is gained when the value of such a variable becomes known" (Anderson & Johnnesson, 2006; Haykin, 1988).

MATB-II involves five tasks: Light, Scale, Tracking (TRACK), Communication (COMM), and Resource Management (RESMAN). For each task, when all decision alternatives are assumed to be equally likely, the information content of a stimulus set, in bits, can be denoted as:

$$H_s(i) = \log_2[j(i)] \qquad (1)$$

where i represents MATB-II tasks (e.g., Light, Scale, TRACK, COMM, and RESMAN). j is the total number of stimuli per set.

To measure the capacity of each task, the baud rate of task i, B(i), is defined as:

$$B(i) = \frac{Hs(i)}{\Delta T(i)} \quad (2)$$

where $H_s(i)$ is the information content of a stimulus set. $\Delta T(i)$ is the duration of task i (time interval between start of one stimulus and start of next occurring stimulus). Then,

$$B(i) = \frac{\log_2[j(i)]}{\Delta T(i)} \quad (3)$$

4.1.1 Baud Rate for SYSMON Task

There are two subtasks included in the SYSMON task window: Lights (upper) and Scales (bottom) (see Figure 2.2).

For the Lights task, each light sends out two signals ("ON" and "OFF") and the operator only needs to respond to one of them (the green light and the presence of the red light). Therefore, there are four decision alternatives in the Lights task, j(L) = 4, and green and red lights are equally weighted (p_L = 0.5). The baud rate for the Lights task with different ΔT are shown below:

$$\Delta T(L) = 10 \text{ sec}, B(L) = \frac{\log_2(4)}{10} = 0.2 \text{ bps} \quad (4)$$

$$\Delta T(L) = 5 \text{ sec}, B(L) = \frac{\log_2(4)}{5} = 0.4 \text{ bps} \quad (5)$$

For the Scales task, there are four moving scales and the operator needs to click on the scales when s/he detects a deviation from the midpoint. Each scale has two directions and there are eight decision alternatives for the Scales task, j(S) = 8, and they are equally weighted (p_S = 0.125). The baud rate for the Scales task with different ΔT are shown below:

$$\Delta T(S) = 30 \text{ sec}, B(S) = \frac{\log_2(8)}{30} = 0.1 \text{ bps} \quad (6)$$

$$\Delta T(S) = 15 \text{ sec}, B(S) = \frac{\log_2(8)}{15} = 0.2 \text{ bps} \quad (7)$$

4.1.2 Baud Rate for COMM Task

During the COMM tasks, audio messages with particular "Call Sign" (e.g., "NASA504") are played and the operator is required to listen to the content and respond by selecting the appropriate radio channel and frequency on the display. There are four channels for the operator to select, j(C) = 4, and they are equally weighted (p_C = 0.25). The baud rate for the COMM task with different ΔT are shown below:

$$\Delta T(C) = 20 \text{ sec}, B(C) = \frac{\log_2(4)}{20} = 0.1 \text{ bps} \quad (8)$$

$$\Delta T(C) = 10 \text{ sec}, B(C) = \frac{\log_2(4)}{10} = 0.2 \text{ bps} \quad (9)$$

4.1.3 Baud Rate for RESMAN Task

During the Resource Management tasks, the operator is required to maintain the level of fuel in tanks A and B within ±500 units of the initial condition of 2500 units each. The acceptable target ranges are shown as a shaded area on the outside of each tank.

In order to maintain the fuel level in tanks A and B, the operator must transfer fuel from the lower supply tanks through the use of the pumps. There are eight pumps and each of them has three states ("ON", "OFF", and "FAILED") and the pump is only operative when it is at the "ON" and "OFF" states. Therefore, each pump will send out three signals and the operator only needs to respond to two of them. There are 16 decision alternatives for the eight pumps in the RESMAN task, j(RM) = 16, and they are equally weighted (p_{RM} = 0.0625). The baud rate for the RESMAN task with different ΔT are shown below:

$$\Delta T(RM) = 40 \text{ sec}, B(RM) = \frac{\log_2(16)}{40} = 0.1 \text{ bps} \quad (10)$$

$$\Delta T(RM) = 20 \text{ sec}, B(RM) = \frac{\log_2(16)}{20} = 0.2 \text{ bps} \quad (11)$$

4.1.4 Baud Rate for TRACK Task

The TRACK task is a perturbation–rejection task and it involves a moving circle which represents the cursor and a fixed circle located in the center of the tracking window which represents the target that the cursor has to track during the task.

The information capacity of the TRACK task can be estimated based on Fitts' law (1954). The index of performance (IP) is analogous to channel capacity in Shannon's theory and IP is calculated by task's index of difficulty (ID) and the movement time (MT):

$$IP = ID/MT \quad (12)$$

$$ID = \log_2\left(\frac{2A}{W}\right) \quad (13)$$

Denote the cursor diameter as D which is half of the effective width (W). Denote the cursor velocity as V. V has three levels in MATB-II: low, medium, and high (Santiago-Espada et al., 2011).

$$V = \frac{dD}{dt} \quad (14)$$

Then, the displacement of the cursor from the target center (Δa) over time ΔT is

$$\Delta a = \frac{\int_0^T V dt}{\Delta T} \quad (15)$$

During the operation, TRACK requires subject to maintain the cursor on target. The information content of TRACK task is:

$$Hs(T) = \log_2\left(\frac{2A}{W}\right) = \log_2[\frac{\Delta a}{D}] \quad (16)$$

Then, the baud rate of TRACK task, B(T), is

$$B(T) = \log_2[\frac{\Delta a}{D}] / \Delta T \quad (17)$$

The cursor velocity has three fixed levels in TRACK task. The values of Δa and ΔT associated with these three levels are (Phillips et al., 2007):

Low: $\Delta a = 3D$, $\Delta T = 6.5$ s

Medium: $\Delta a = 3D$, $\Delta T = 2.3$ s.

High: $\Delta a = 3D$, $\Delta T = 1.7$ s.

Then, the baud rates, B(T), of TRACK task are:

$$\text{Low: } B(T) = \frac{\log_2(3D/D)}{6.5} = 0.24 \text{ bps} \quad (18)$$

$$\text{Medium: } B(T) = \frac{\log_2(3D/D)}{2.3} = 0.69 \text{ bps} \quad (19)$$

$$\text{High: } B(T) = \frac{\log_2(3D/D)}{1.7} = 0.93 \text{ bps} \quad (20)$$

4.2 Task Weight and Overall Baud Rate

The overall baud rate for MATB-II tasks is the summation of the baud rate from all subtasks, denoted as B_{TOT}. Then,

$$B_{TOT} = \sum B(i) \tag{21}$$

A task weight is assigned to each MATB-II subtask, denoted as w(i) for subtask i. It is based on the baud rate for task i and the overall baud rate of the system. Then,

$$w(i) = \frac{B(i)}{B_{TOT}} \tag{22}$$

Levels of task difficulty of the system are determined by different levels of $B_{TOT.}$

To assess user performance, the operator's responses are recorded through the system. During the MATB-II operation, the human operator generates an output baud rate for each MATB-II task, denoted as $B_H(i)$. The human output baud rate for each task is based on the response ratio for task i, RR(i). The response ratio is defined as the correct responses per trial for subtask i and denoted as RR(i).

$$RR(i) = \frac{\text{Number of Correct Responses of Subtask i in a Trial}}{\text{Total Number of Events of Subtask i in a Trial}} \tag{23}$$

Thus, a human output baud rate for task i is

$$B_H(i) = RR(i)B(i) \tag{24}$$

The overall human output baud rate for the MATB-II operation is denoted as B_{HTOT}.

$$B_{HTOT} = \sum B_H(i) \tag{25}$$

A human output task weight, $w_H(i)$, is calculated based on the human output baud rate for task i and the overall human output baud rate. Then,

$$w_H(i) = \frac{B_H(i)}{B_{HTOT}} \tag{26}$$

4.3 Rearrangement of Task Weights

To overcome the limitation of baud rate selection (unbalanced between tasks) and improve overall multitasking performance for each human operator, this section introduces an approach to rearrange task weights for each participant.

The objective of the rearrangement approach is to maximize the overall human output baud rate with a fixed amount of the overall baud rate from MATB-II system. Thus, the objective function is B_{HTOT} with a new set of task weights, w'(i).

From the previous section, the objective function is equivalent to $B_{TOT}\sum RR(i)w'(i)$, because

$$B_{HTOT} = \sum B_H(i) = \sum RR(i)B(i) \tag{27}$$

$$B(i) = w'(i)B_{TOT} \tag{28}$$

$$B_{HTOT} = \sum RR(i)B(i) = \sum RR(i)w'(i)B_{TOT} = B_{TOT}\sum RR(i)w'(i) \tag{29}$$

In the objective function, the B_{TOT} is a constant. The goal of this approach is to find a set of solutions that maximize the summation of $B_{TOT}\sum RR(i)w'(i)$.

To achieve this goal, a cross-entropy optimization approach with cross-entropic constraints is applied (Fang, Rajasekera, & Tsao, 1997, pp. 150-151). In this setting, the objective of this approach is to find a probability distribution, w'(i), that is the "closest" to a given a priori distribution, w(i), within a "deviation measure" of ε from a given set of distributions. This problem can be descripted as following forms:

$$\text{Maximize} \quad B_{TOT} \sum RR'(i) w'(i) \tag{30}$$

$$\text{Subject to} \quad \sum w'(i) = 1$$

$$RR'(i) = RR(i)$$

$$0 < w'(i) < 1$$

$$| \text{CrossEnt}[w(i), w'(i)] | \leq \mathcal{E}$$

Where CrossEnt[w(i), w'(i)] is the cross-entropy of task weight and

$$\text{CrossEnt}[w(i), w'(i)] = -\sum w(i) \ln \frac{w'(i)}{w(i)} \tag{31}$$

In this setting, there are three sets of unknown variables, RR(i), w(i), and $\mathcal{E}$. Several assumptions based on the operation of MATB-II tasks are made: (1) the response ratio, RR(i), for task i is a constant for each human operator and the values of RR(i) are obtained from the operation of a set of equally weighted MATB-II tasks (this assumption is justified in section 7.1.2); (2) the human operator needs to pay attention to all MATB-II tasks during the operation, then, the new task weight for task i, w'(i), cannot be equal to 0 or 1; (3) the "deviation measure" of $\mathcal{E}$ represents the system allowance and is defined by the system requirement in this study. A constant value is set up for $\mathcal{E}$ (= 0.1) in this study since operators are expected to pay attention to all MATB-II tasks across all task conditions in this study. It promotes and motivates the equal weighting among different subtasks from human operator during the operation in the system. An example will demonstrate the calculation of the cross-entropy of task weight and task weights rearrangement in Chapter 6 (see section 6.2.5.1).

This optimization problem can be solved with an optimization tool (e.g., LINGO 13.0) and one or more sets of solutions can be obtained to achieve an optimal value for $B_{TOT}\sum RR(i)w(i)$ or B_{HTOT}.

In this study, a new set of task weights, w'(i), was selected from the solutions for each human operator. A modified set of task baud rates, B'(i), was assigned to each operator according to $B'(i) = B_{TOT}w'(i)$. They were asked to perform the modified set of MATB-II tasks and it was expected that there would be an improvement of user performance after this rearrangement.

In MATB-II tasks, the baud rate for TRACK task was three fixed values defined by the MATB-II system. Thus, B(T)'s were three constant values in calculation for all three levels of task difficulty in this study.

5 Phase I Experiment Methodology

5.1 Objectives and Rationale

To validate the proposed quantitative model, first, a performance baseline needed to be obtained from equally weighted multitasks. Phase I was intended to apply the proposed model and identify a performance baseline for each individual in a multitasking environment, MATB-II.

To determine the performance baseline, a set of equally weighted multitasks was completed by each participant during Phase I. Then, Phase II followed the proposed approach to rearrange task weights for each participant and to investigate the improvement of user performance in MATB-II tasks. However, as mentioned in the literature review, practice and training will enhance user performance in multitasking. To eliminate the carryover effect from training, a practice session was performed by each participant and the participant was invited to the following experiment sessions if s/he met a set of training criteria.

At the end of this study, comparisons of user performance were performed between Phase II and Phase I experiments to determine the improvement between Phase II and Phase I.

Another objective of Phase I was to investigate the effects of WMC and task difficulty on multitasking performance. To evaluate the effect of WMC on multitasking performance, a WMC test session was completed at the initial contact with potential participants to estimate the levels of their WMC. To determine the effect of task difficulty in multitasking, three levels of task difficulty were selected and applied in both practice and experimental sessions.

5.2 Methods

5.2.1 Participants

Twenty-five participants (11 male, 14 female, average age = 19.64, SD = 1.85) were recruited from university for this study. All participants were required to be in good health with normal hearing and normal or corrected-to-normal vision. Only right-handed participants were recruited for this study due to the operation of MATB-II and to avoid the influence of handedness on the experiment. In addition, English was required as native language or primary language of all participants. The health state, vision, handedness, and language of potential participants were confirmed upon initial contact (a demographic questionnaire and handedness survey used in this study are provided in the Appendices). The handedness survey was based on the Edinburgh Handedness Inventory (Oldfield, 1971). All participants were required to have no previous experience with MATB-II or similar training.

In addition, the WMC of potential participants was estimated by a WMC test (see section 5.2.2). Participants who completed the WMC test were recruited for this research.

5.2.2 WMC Test

WMC has been reported as a predictor of multitasking performance and can be predicted by working memory span tasks (Kane & Engle, 2003; König et al., 2005; Redick & Engle, 2006; Bühner et al., 2006). For instance, the operation span task (OSPAN) (Turner & Engle, 1989) has been established as a reliable and valid marker of WMC. An automated version of OSPAN was employed in this study (Unsworth et al., 2005) since it is easy-to-administer and requires little intervention from experimenter that participants can complete it online by themselves.

The automated OSPAN is a computerized test that is mouse driven and allows the participant to complete the task independently of the experimenter. The test includes items (letters) to remember and math problem to solve (Unsworth et al., 2005). There are three practice sessions and one experimental session. Appendix C includes an instruction of the automated OSPAN task and it was provided to participants during the experiment.

The score range of this automated OSPAN test was 0 – 75. At the end of the test, the program reported five scores: OSPAN score, total number correct, math errors, speed errors, and accuracy errors (Unsworth et al., 2005). The test required approximately 20–25 min to complete for each participant. Figure 5.1 shows the procedure of OSPAN test.

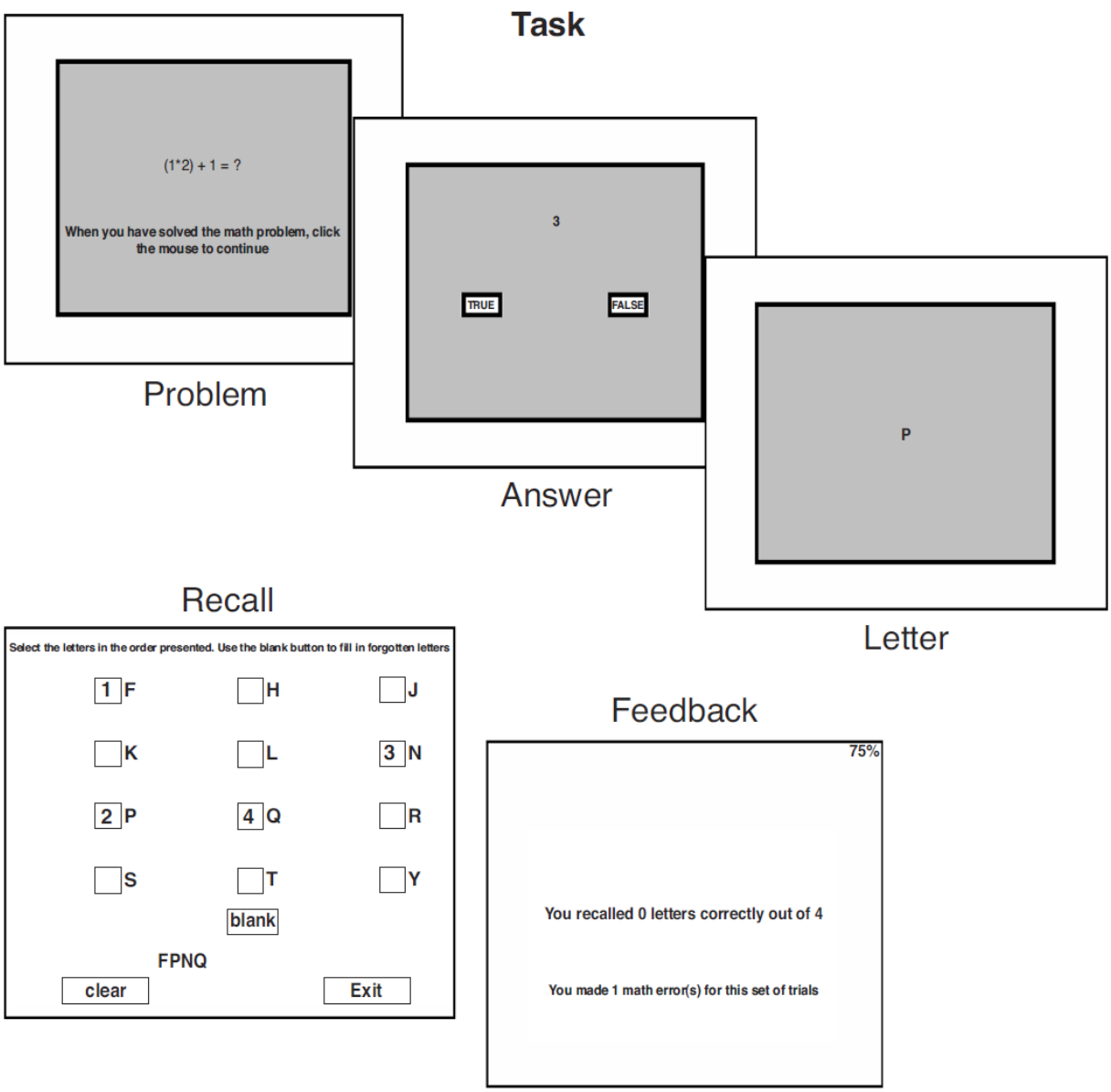


Figure 5.1 Illustration of the automated operation span test. (Unsworth et al., 2005)

OSPAN scores also remain stable over test–retest intervals of a few minutes (r = .77–.79; Turley-Ames & Whitfield, 2002), to 3 weeks (stability coefficient = .82; Klein & Fiss, 1999), to 3 months (stability coefficient = .76; Klein & Fiss, 1999). Experimental tasks (e.g., MATB-II tasks) can therefore follow OSPAN by as little as a few minutes, and as many as 3 months.

5.2.3 Apparatus and Materials

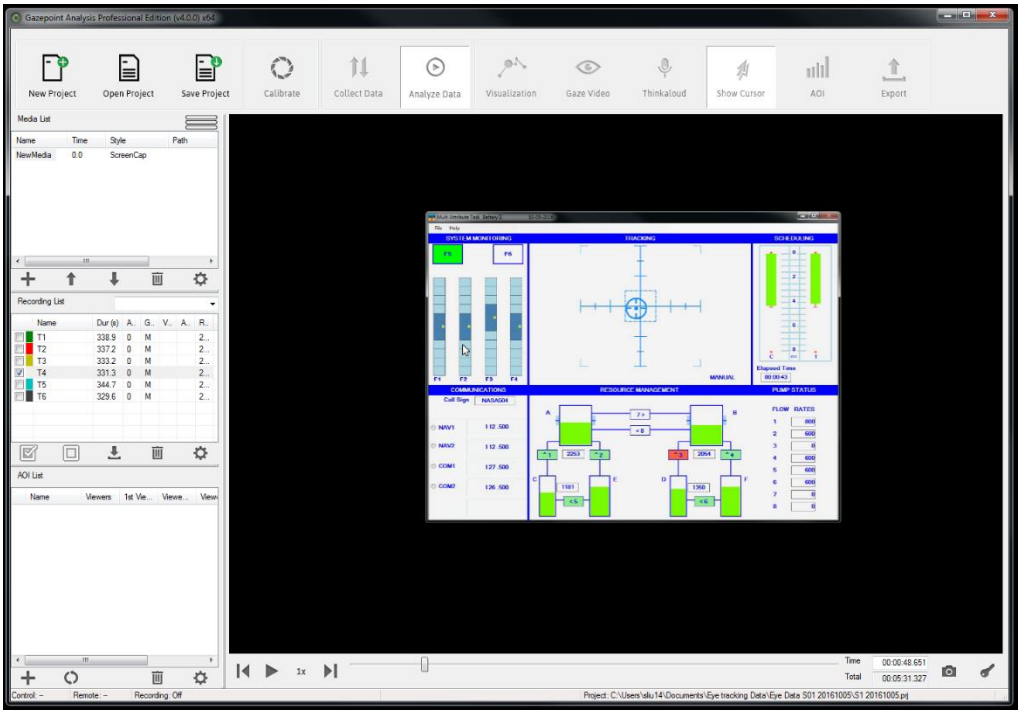
Before the experiment, a demographic questionnaire and a handedness survey were provided to each participant (see Appendix A and B). Each participant was required to complete an automated OSPAN test to evaluate their WMC.

The Multi-Attribute Task Battery II (MATB-II) was employed in this study as a platform to assess multitasking. Five MATB-II tasks were tested in this study: Light, Scale, Tracking (TRACK), Communication (COMM), and Resource Management (RESMAN). For the Light task, participants needed to make a response to the absence of the green light and the presence of the red light. For the Scale task, participants needed to monitor the four moving scales and click on the bars when they detected a deviation from the midpoint. For the COMM task, audio messages with particular “call sign” were played and participants were required to listen to the content and respond by selecting the appropriate radio channel and frequency on the display. A joystick is required for the tracking task in MATB-II (Santiago-espada et al., 2011). When the TRACK task was under manual mode, participants were asked to use a joystick with their left hands to keep the circular target in the center of the tracking window (the non-dominant hand was selected since tracking task was not a primary task in this study (Phillips, Kinsler, Repperger et al., 2013; Rupp et al., 2014)). For the RESMAN task, participants were asked to maintain the level of fuel in tanks

A and B within ±500 units of the initial condition of 2500 units each. All participants were required to use a mouse with their right hands (dominant hands) to make response to Light, Scale, COMM, and RESMAN tasks. A NASA-TLX questionnaire was presented after each trial and a NASA-TLX pairwise comparisons of factors form was completed by each participant at the end of the experiment (see Appendix E).

Physiological responses, eye movement measures, were recorded during the experiment for each individual. The device used in this study was the Gazepoint eye tracking system. Figure 5.2 shows the hardware and software of the eye tacking system. The eye tracking system recorded eye movement data at 60 Hz with an accuracy of 0.5 – 1 degree of visual angle. It can capture 25cm x 11cm (horizontal x vertical) movement and ±15 cm range of depth movement. Therefore, participants were required to keep their body in a small range and look at the monitor during the experimental trials. Data recorded by the eye tracking system for this study included blink rate and blink duration (details of eye measures were introduced in following sections). After the experiment, eye measures of each participant were processed using the software, Gazepoint Analysis. Figure 5.2 (b) shows the screenshot of Gazepoint Analysis with one experimental trial data. After selected desired experimental trials in the Recording List (middle left), the software can export processed eye measures by clicking “Export” button on the upper-right corner. Then, processed eye measures were saved in CSV files for each experimental trial.

(a) Gazepoint Eye Tracker



(b) Gazepoint Analysis

Figure 5.2 Hardware and software of eye tracking system

5.2.4 Independent Variables

There were two independent variables manipulated in Phase I: task difficulty and WMC.

5.2.4.1 Task difficulty

Task difficulty was manipulated by different levels of overall baud rate, B_{TOT}, of MATB-II tasks. There were three levels of task difficulty in this Phase: low, medium, and high. Each level of task difficulty was manipulated through the overall baud rate, B_{TOT}. For each individual task in MATB-II, the baud rate was calculated based on the time interval between two events in task i and equations of section 4.1.

To determine an appropriate range of task difficulty, a pilot test manipulated the B_{TOT} to select the task difficulty levels (Liu et al., 2016). The study included four MATB-II tasks (Lights, Scales, COMM, and TRACK) and three levels of task difficulty (B_{TOT} = 0.48, 1.04, and 1.63 bps for low, medium, and high levels, respectively). These values of B_{TOT} were selected from the pilot test in the previous study and listed in Table 5.1. The time interval of task i, $\Delta T(i)$, and the baud rate of task i, B(i), are also listed in Table 5.1. The baud rate of TRACK task, B(T), was predefined by MATB-II. It had three levels with a fixed value for each level.

Table 5.1 Baud Rate and Task Difficulty in Pilot Test

Task Difficulty		Low	Medium	High
Time interval (second)	$\Delta T(L)$	13	9	4
	$\Delta T(S)$	39	25	13
	$\Delta T(C)$	26	17	9
Baud Rate (bps)	B(L)	0.15	0.22	0.5
	B(S)	0.08	0.12	0.23
	B(C)	0.08	0.12	0.22
	B(T)	0.24	0.69	0.93
	B_{TOT}	0.55	1.15	1.88

The pilot test included ten subjects and the results of user performance in MATB-II showed a relatively high response ratio (RR) (> 91%) for all three decision making tasks (e.g., Lights, Scales, and COMM). For the purpose of the current study, higher baud rates than those used in the pilot test was desired, therefore additional pilot trials were performed to adjust the baud rates for each task. A series of several MATB-II tasks with higher overall baud rates were completed by two subjects, and the appropriate overall baud rate for Phase I was selected when the overall RR lower than 65% was detected at medium level of task difficulty. Table 5.2 shows the adjusted baud rate and task difficulty for Phase I.

Table 5.2 Baud Rate and Task Difficulty in Phase I

Task Difficulty		Low	Medium	High
Time interval (second)	ΔT(L)	20	10	5
	ΔT(S)	30	15	7
	ΔT(C)	20	10	5
	ΔT(RM)	40	20	10
Baud Rate (bps)	B(L)	0.1	0.2	0.4
	B(S)	0.1	0.2	0.43
	B(C)	0.1	0.2	0.4
	B(RM)	0.1	0.2	0.4
	B(T)	0.24	0.69	0.93
	B_{TOT}	0.64	1.49	2.56

5.2.4.2 WMC

WMC was assessed by the automated OSPAN test. As described in Section 5.2.2, the potential OSPAN scores ranged from 0 – 75 of the WMC test, participants' OSPAN scores covered a relative wide range (28 – 75).

Table 5.3 shows a summary of the results from the OSPAN test. Only the OSPAN scores were applied to the data analysis as an indicator of WMC for the participants.

Table 5.3 Summary of OSPAN Test

	OSPAN Score	Total correct	Math error	Speed error	Accuracy error
Mean	53.42	66.62	2.38	0.69	1.69
SD	12.28	5.73	2.02	0.93	1.49
Min	28	56	0	0	0
Max	75	75	8	4	7

5.2.5 Dependent Variables

To validate the proposed model and investigate the effects of task difficulty and WMC on multitasking performance, Phase I used several dependent measures, which can be categorized into three types of variables: MATB-II performance, Workload Rating Scale (WRS), and eye movement measures.

5.2.5.1 MATB-II Performance

MATB-II performance included average response time (RT) for Light, Scale, and COMM tasks, response ratio (correct response per trial, RR) for Light, Scale, COMM, and RESMAN tasks, as well as root mean square deviation (RMSD) for TRACK task.

Average response time (RT(i)). When the stimulus of an event was first triggered, RT(i) was the average time taken for the participant to respond to each event of the MATB-II subtask i within each trial. The index i represented each MATB-II subtask.

Response Ratio (RR(i)). RR(i) was the percentage of correct responses of MATB-II subtask i during each experimental trial. The index i represented each MATB-II subtask.

Root mean square deviation (RMSD). RMSD from center point was calculated and recorded along with the number of samples during a predefined interval for TRACK task. RMSD was in pixel units.

5.2.5.2 Workload Rating Scale (WRS)

At the end of each experimental trial, a WRS was presented to the participant. It was a NASA-TLX rating scale and included six parts: mental demand (MD), physical demand (PD), temporal demand (TD), the individual's perceived level of performance (PE), effort (EF), and frustration (FR). Appendix C shows an instruction of WRS.

A NASA-TLX pairwise comparisons of factors form (Hart & Staveland, 1988) was provided at the end of MATB-II experiment session of both Phase I and Phase II for each participant. The weighted NASA-TLX rating score was calculated and used as an indicator of WRS.

5.2.5.3 Eye Movement Measures

The physiological response measures included blink rate and blink duration in this study. Blink rate is the number of blinks per minute (blinks/minutes). Blink duration is the duration of the preceding blink in seconds.

These eye movement measures were recorded through the eye tracking system, Gazepoint, during the experiment. Gazepoint Analysis is the software to integrate the eye movement data recorded by Gazepoint Control during the experiment. In order to account for individual differences in eye measures, a baseline trial was collected for each participant. The physiological

responses from the eye tracking system (blink rate and blink duration) were normalized based on the baseline measures for each participant by using the following equations:

$$\text{Normalized Blink Rate (\%BKR)} = \frac{Observed\ Blink\ Rate - Individual\ Blink\ Rate\ Baseline}{Individual\ Blink\ Rate\ Baseline}$$

$$\text{Normalized Blink Duration (\%BKD)} = \frac{Observed\ Blink\ Duration - Individual\ Blink\ Duration\ Baseline}{Individual\ Blink\ Duration\ Baseline}$$

5.2.6 Procedure

Upon initial contact with the potential participants, an online screening session was required, including a demographic questionnaire and a handedness survey to verify their age, native language, health condition, and handedness (see Appendix A and B). If they were qualified, an online WMC test (the automated OSPAN test) was required to complete by themselves before Phase I experiment. An instruction of the automated OSPAN test was provided to help them to complete the WMC test (see Appendix C).

Phase I experiments were conducted in the Cognitive Lab (Daniels Hall, North Carolina State University). When the qualified participants arrived, an informed consent form was provided. After s/he completed the informed consent form, an instruction of the experiment was provided including an introduction of the procedure of two phases, an instruction of MATB-II tasks, and an instruction of eye tracking system. Then, a calibration of the eye tracking system was completed.

During Phase I, a two-hour practice sessions of multitasks in MATB-II was required for each participant before the experimental session. Each participant completed twelve 5-minute trials of MATB-II tasks for practice. The twelve trials covered three levels of task difficulty and there were four replicates in each level. The order of these practice trials were randomized for each

participant. User performance were recorded during practice. At the end of the practice session, if the participant had an average RR of 65% or higher at the medium level of task difficulty trials, s/he was invited to participate in Phase I and Phase II experimental sessions (Bowers et al., 2014). After the practice session, a 30-minute break was provided to the participants before the MATB-II experimental session. Then, to obtain a baseline for their physiological responses, participants were asked to look at the screen and sit as naturally as possible for 5 minutes. A static window of MATB-II platform was presented on the screen during the baseline trial. Figure 5.3 shows an example of the baseline trial.

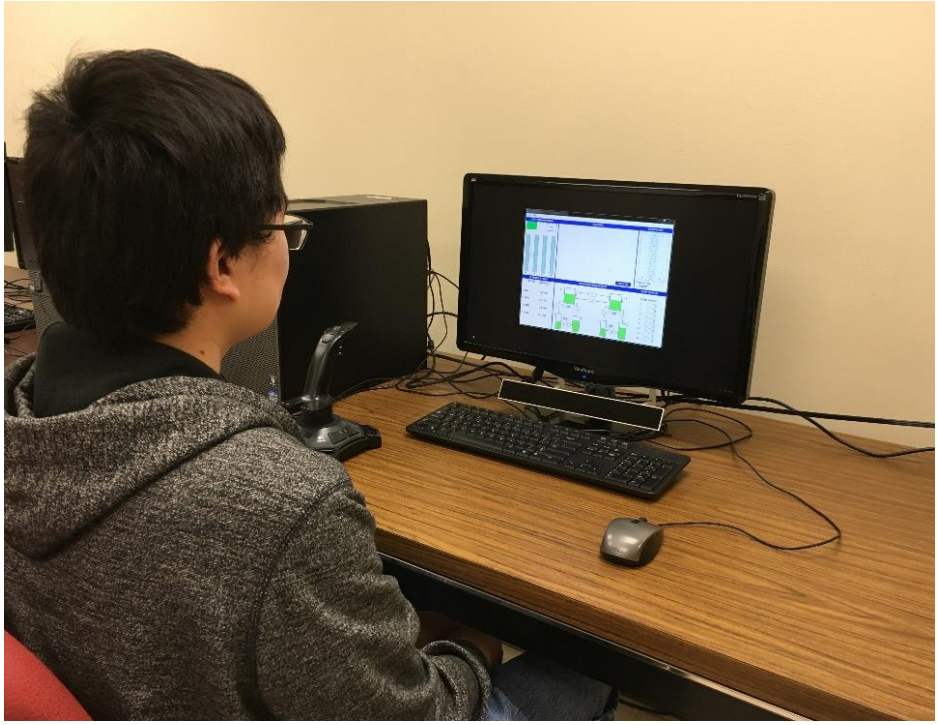


Figure 5.3 Baseline trial

The MATB-II experimental session was comprised of six 5-minute MATB-II trials. There were two trials in each level of task difficulty. During the operation of MATB-II tasks, participants were asked to complete each MATB-II subtask as quickly and as accurate as possible by switching between different subtasks. Their MATB-II performance, NASA-TLX rating scores, and eye movement measures were recorded during the experimental session. At the end of the experiment, a NASA-TLX pairwise comparisons of factors form (Hart & Staveland, 1988) was completed by the participant (see Appendix E). Figure 5.4 shows the procedure of Phase I.

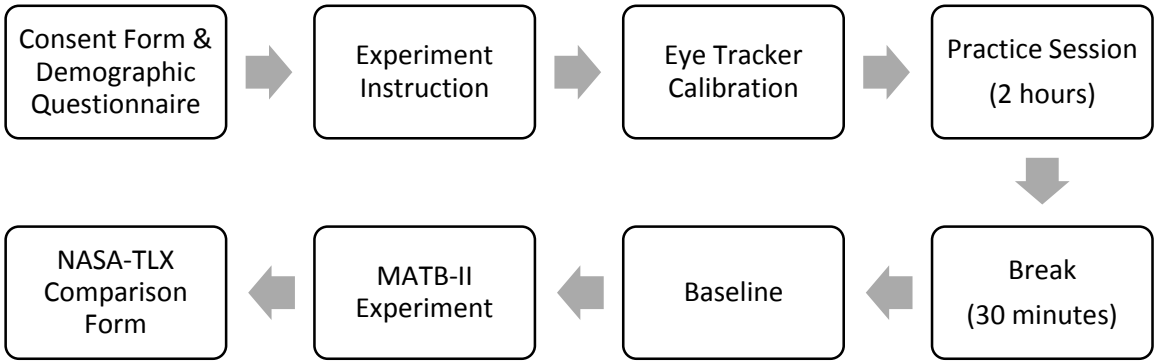


Figure 5.4 Phase I Procedure

5.2.7 Experiment Design and Statistical Data Analysis

To investigate the effects of task difficulty and WMC on multitasking performance, analysis of variance (ANOVA) was applied. It was a split-plot design with participants as whole plot and experimental trials as split plots. The treatment factors included WMC as the whole plot

factor and task difficulty as the split plot factor. The model of the experiment design in Phase I is as follows:

$$y_{ijk} = \mu + \alpha_i + \varepsilon\,(W)_{ij} + \beta_k + (\alpha\beta)_{ik} + \varepsilon(S)_{ijk}$$

where:

y_{ijk} = response variable (response ratio (RR), response time (RT), RMSD in MATB-II; WRS score; normalized blink rate (%BKR), normalized blink duration (%BKD));

μ = response mean;

α_i = whole plot main effect, WMC, $i \in [0, 75]$;

$\varepsilon\,(W)_{ij}$ = whole plot error, j = participant number; $\varepsilon\,(W)_{ij} \sim^{iid} \sim N(0, \sigma_W^2)$.

β_k = split plot main effect, task difficulty, j = 1, 2, 3;

$(\alpha\beta)_{ik}$ = whole plot and split plot interaction;

$\varepsilon(S)_{ijk}$ = split plot error; $\varepsilon(S)_{ijk} \sim^{iid} \sim N(0, \sigma_S^2)$.

Before perform ANOVA, the assumptions of homoscedasticity and residual normality were examined using Bartlett's test (Garson, 2012) and Shapiro-Wilk normality test (Shapiro and Wilk, 1965), respectively. For any data sets that did not satisfy the assumptions of homoscedasticity or residual normality, transformations were applied to the response variables (e.g., log transformation). Given the transformations were ineffective, the average ranks of response variables were applied as an alternative to a nonparametric analysis.

None of the response variables or their transformations satisfied the homoscedasticity and residual normality assumptions. Since the results using the average ranks produced similar results to ANOVAs using the original measures for all response variables, analyses using original measures were considered valid (Montgomery, 1997) and were performed and are reported in the following section.

5.3 User Performance Baseline

At the end of the practice session for each participant, the average RR of MATB-II tasks at medium level of task difficulty was calculated and the participant with an average RR of 65% or higher was invited to the MATB-II experiment session. During this study, all of the 25 participants had an average RR of 65% or higher during their practice sessions and were invited to the MATB-II experiments sessions of both Phase I and Phase II.

After Phase I experiment session, participants' task performance was evaluated and summarized as their performance baseline. Table 5.4 shows a statistical summary (mean and SD) of MATB-II task performance in Phase I by task difficulty levels. The response metrics included RT (second) of Light, Scale, and COMM tasks (RT_Light, RT_Scale, and RT_COMM), RR (ratio) of Light, Scale, COMM, and RMAN tasks (RR_Light, RR_Scale, RR_COMM, and RR_RMAN), and RMSD (pixel) of TRACK task.

Table 5.4 Summary of MATB-II Performance Baseline (Mean (SD))

Task Difficulty	Low	Medium	High
RT_Light	2.28 (0.64)	2.70 (0.81)	2.59 (0.62)
RT_Scale	4.01 (1.39)	4.40 (1.43)	5.25 (1.36)
RT_COMM	1.55 (0.53)	1.71 (0.55)	2.18 (0.46)
RR_Light	0.99 (0.02)	0.99 (0.02)	0.92 (0.05)
RR_Scale	0.82 (0.28)	0.85 (0.23)	0.88 (0.16)
RR_COMM	0.93 (0.08)	0.93 (0.08)	0.85 (0.11)
RR_RMAN	0.89 (0.09)	0.89 (0.08)	0.86 (0.10)
RMSD	18.93 (7.95)	30.80 (12.03)	47.90 (19.61)

Subjective workload was assessed by the NASA-TLX rating scale and physiological responses, including blink rate and blink duration, were recorded through the eye tracking system during the experiment. Table 5.5 shows a statistical summary (mean and SD) of subjective workload (WRS) and eye measures (%BKR and %BKD) in Phase I by task difficulty levels. The WRS results showed the weighted subjective rating scores from NASA-TLX rating scales and a higher score indicated a higher level of mental workload. The eye measures, %BKD and %BKR, were normalized blink duration and blink rate which represented the percentage of changes in these eye measures from baseline trial for each participant.

Table 5.5 Summary of Workload Baseline (Mean (SD))

Task Difficulty	Low	Medium	High
WRS	40.29 (18.69)	49.54 (17.16)	65.42 (15.84)
%BKD	-12.31% (11.17%)	-14.01% (11.94%)	-10.92% (21.98%)
%BKR	-44.20% (32.97%)	-44.36% (25.79%)	-52.83% (32.29%)

From Table 5.5, a general trend of increased mental workload with the increase of task difficulty was detected by the increase of WRS.

6 Phase II Experiment Methodology

6.1 Objectives and Rationale

Phase II was intended to validate the proposed model using a rearranged set of multitasks for each individual. Also, Phase II aimed at investigating the effects of WMC and task difficulty on multitasking performance.

With the baseline performance from the first phase, a new set of multitasks was designed for each participant according to the rearrangement of task weights determined from the proposed model. The same participants who completed all sessions in Phase I were invited to the experiment in Phase II. During the experimental session, each participant completed a rearranged set of multitasks which was customized for each individual. After the experiment, comparisons of user performance were performed between Phase II and Phase I experiments to validate the proposed model.

The new set of multitasks covered all three levels of task difficulty, the same as in the first phase, to determine the effect of task difficulty on multitasking performance. The effect of WMC on multitasking performance was also investigated and the WMC scores from Phase I were applied to the data analyses in Phase II.

6.2 Methods

6.2.1 Participants

Participants who met the practice criteria and completed the MATB-II experimental session in Phase I were recruited for Phase II. All 25 participants from Phase I were invited and completed Phase II.

6.2.2 Apparatus and Materials

The Multi-Attribute Task Battery II (MATB-II) was also employed in this phase. A rearranged set of multitasks was designed for each participant. Five tasks were included in the experiments: Light, Scale, Tracking (TRACK), Communication (COMM), and Resource Management (RESMAN). A desktop, a mouse, and a joystick were provided for the MATB-II operation. All participants were required to use a joystick with their left hand to complete the TRACK task and use a mouse with their right hand to respond to the Light, Scale, COMM, and RESMAN tasks. A NASA-TLX questionnaire was presented after each trial and a NASA-TLX pairwise comparisons of factors form was completed by each participant at the end of the experiment.

An eye tracking system, Gazepoint, collected eye movement measures for each participant during the experimental trials in Phase II. Data recorded by the eye tracking system included blink rate and blink duration.

6.2.3 Independent Variables

There were two independent variables applied in Phase II: task difficulty and WMC.

6.2.3.1 Task difficulty

There were three levels of task difficulty in this Phase: low, medium, and high. The overall baud rate for each level of task difficulty was the same as the one in Phase I (see Table 5.1). During Phase II, task weights were rearranged according to the baseline performance for each individual in Phase I. Though the baud rate for each MATB-II subtask was adjusted, the overall baud rate, B_{TOT}, maintained the same value for each level of task difficulty in Phase I (e.g., B_{TOT} = 0.64, 1.49, and 2.49 bps, for low, medium, and high level, respectively).

6.2.3.2 WMC

WMC had been assessed by the automated OSPAN test in Phase I. As described in Section 5.2.2 and 5.2.4, the OSPAN scores was the indicator of WMC and ranged from 0 – 75. A retest of OSPAN test was not necessary for Phase II since all participants were from Phase I and the OSPAN scores will remain stable over test-retest intervals of a few minutes to 3 months (Klein & Fiss, 1999).

6.2.4 Dependent Variables

To validate the proposed model, evaluate the performance improvement, and investigate the effects of task difficulty and WMC on multitasking performance, Phase II used the same dependent variables as in Phase I. These measures included three types of variables: MATB-II performance, Workload Rating Scale (WRS), and eye movement measures.

6.2.4.1 MATB-II Performance

MATB-II performance included RTs for Light, Scale, and COMM tasks, RRs for Light, Scale, COMM, and RESMAN tasks, as well as RMSD for TRACK task. These measures were recorded during the MATB-II experiment.

6.2.4.2 Workload Rating Scale (WRS)

At the end of each experimental trial, a WRS, which was a NASA-TLX rating scale, was presented to the participant. A NASA-TLX pairwise comparisons of factors form (Hart & Staveland, 1988) was provided at the end of MATB-II experiment session for each participant. The weighted NASA-TLX rating score was calculated and used as an indicator of WRS.

6.2.4.3 Eye Movement Measures

As same as in Phase I, eye movement measures included blink rate and blink duration in Phase II. The normalization procedure of eye movement measures was the same as described in Phase I. The normalized blink rate (%BKR) and normalized blink duration (%BKD) were used as the eye responses in this study.

6.2.5 Procedure

For all selected participants, Phase II was scheduled after three weeks of their completion of Phase I experiments. There were three sessions included in Phase II. Figure 6.1 shows the procedure of Phase II.

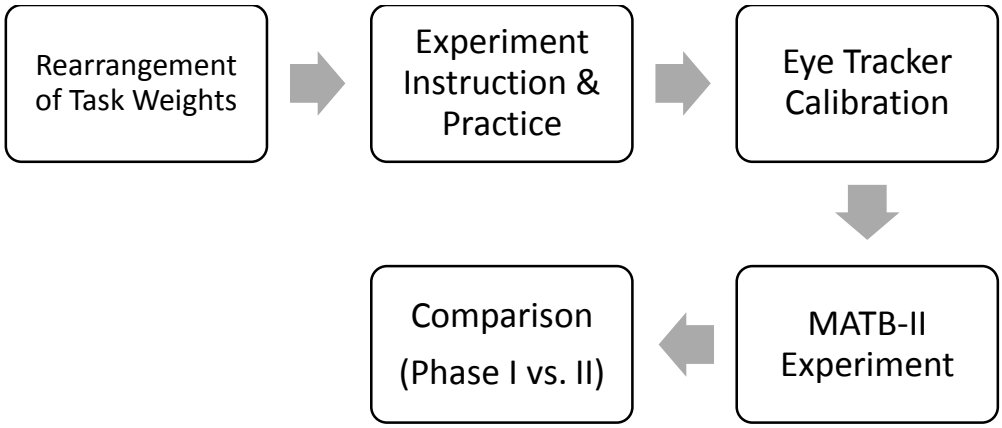


Figure 6.1 Phase II Procedure

6.2.5.1 Rearrangement of Task Weights

After Phase I, user performance was recorded and a rearranged set of MATB-II tasks was provided for each participant before Phase II experiment. The rearrangement of task weights followed the approach which was proposed in the proposed model (see section 4.2).

Table 6.1 shows an example of the rearrangement of task weights for one participant across all three levels of task difficulty. TRACK task had a fixed baud rate at each level of task difficulty, so the rearrangement of task weights was only performed for Light (L), Scale (S), COMM (C), and RMAN (RM) tasks.

Table 6.1 Example of Task Weights Rearrangement

		Task Difficulty		
		Low	**Medium**	**High**
Initial Task Weights (Phase I)	**w(L)**	0.156	0.134	0.156
	w(S)	0.156	0.134	0.168
	w(C)	0.156	0.134	0.156
	w(RM)	0.156	0.134	0.156
	w(T)	0.376	0.464	0.363
Performance Baseline (RR)	**RR(L)**	1	0.98	0.96
	RR(S)	1	0.97	0.96
	RR(C)	0.95	0.93	0.77
	RR(RM)	0.93	0.91	0.88
Rearranged Task Weights (Phase II)	**w'(L)**	0.176	0.160	0.166
	w'(S)	0.176	0.151	0.187
	w'(C)	0.138	0.113	0.134
	w'(RM)	0.134	0.112	0.150
	w'(T)	0.376	0.464	0.363

Note: w(i) = initial task weight of task i; w'(i) = rearranged task weight of task i.

During the rearrangement of task weights, the summation of task weights for Light (L), Scale (S), COMM (C), and RMAN (RM) tasks was 0.625, 0.537, and 0.637 at low, medium, and high level, respectively. For the participant shown in Table 6.1, the results of the cross-entropy of task weight, CrossEnt[w(i), w'(i)], for each task in MATB-II across all three levels of task difficulty are shown in Table 6.2. The range of the cross-entropy of task weight for this participant was [-0.094, 0.098].

Table 6.2 Results of the Cross-Entropy of Task Weight

		Task Difficulty		
		Low	**Medium**	**High**
CrossEnt[w(i), w'(i)]	**L**	-0.074	-0.094	-0.039
	S	-0.074	-0.063	-0.068
	C	0.078	0.092	0.098
	RM	0.096	0.097	0.026

Note: $\text{CrossEnt}[w(i), w'(i)] = -\sum w(i) \ln\frac{w'(i)}{w(i)}$

6.2.5.2 MATB-II Experiment

During the experiment, participants were required to complete the rearranged set of multitasks at each level of task difficulty, twice. Therefore, there were six 5-minute trials of the experiment in Phase II for each participant. During the operation of MATB-II tasks, participants were asked to complete each MATB-II subtask as quickly and as accurately as possible by switching between these subtasks. Their performance and eye responses were recorded during the experiment.

6.2.5.3 Comparison

Users performance in the Phase II experiment were compared with the data from the Phase I experiment to validate the proposed model. Procedure of the comparison was introduced in the next section.

6.2.6 Experiment Design and Statistical Data Analysis

To investigate the effects of task difficulty and WMC on multitasking performance, analysis of variance (ANOVA) was applied. The same as Phase I, Phase II was a split-plot design. The model used for the experiment design in Phase II was as the same as the one used in Phase I (see section 5.2.7). The response variables included response ratio (RR), response time (RT), RMSD in MATB-II; WRS scores; normalized blink rate (%BKR) and normalized blink duration (%BKD).

The first step is to examine the assumptions of homoscedasticity and residual normality using Bartlett's test (Garson, 2012) and Shapiro-Wilk normality test (Shapiro and Wilk, 1965), respectively. None of the response variables or their transformations satisfied the homoscedasticity

and residual normality assumptions. As in Phase I, the results using the average ranks had the similar results of ANOVAs using the original measures for all response variables. Therefore, analyses using original measures were considered valid (Montgomery, 1997) and were performed and reported in the following sections.

To validate the improvement of user performance after the rearrangement of task weights, paired comparisons were conducted. Before performing the paired comparisons, the assumptions of homoscedasticity and residual normality for a t-test needed to be examined. For any data sets that did not satisfy the assumptions, transformations were applied to the response variables (e.g., log transformation). This approach was the same as the approach applied before ANOVA. Since none of the response variables or their transformations satisfied the homoscedasticity and residual normality assumptions, a non-parametric method, Kruskal-Wallis Test (Higgins, 2003), was applied to the paired comparisons. Since the results of significance from Kruskal-Wallis Test were similar to the results from the t-test using original measures for all response variables, the results from the t-test are reported in section 7.1.

A set of tasks was provided in the single multitasking system, MATB-II. To verify the cognitive relations among these tasks in MATB-II, correlation analysis was conducted for all task performance measures and the results are reported in section 7.3.

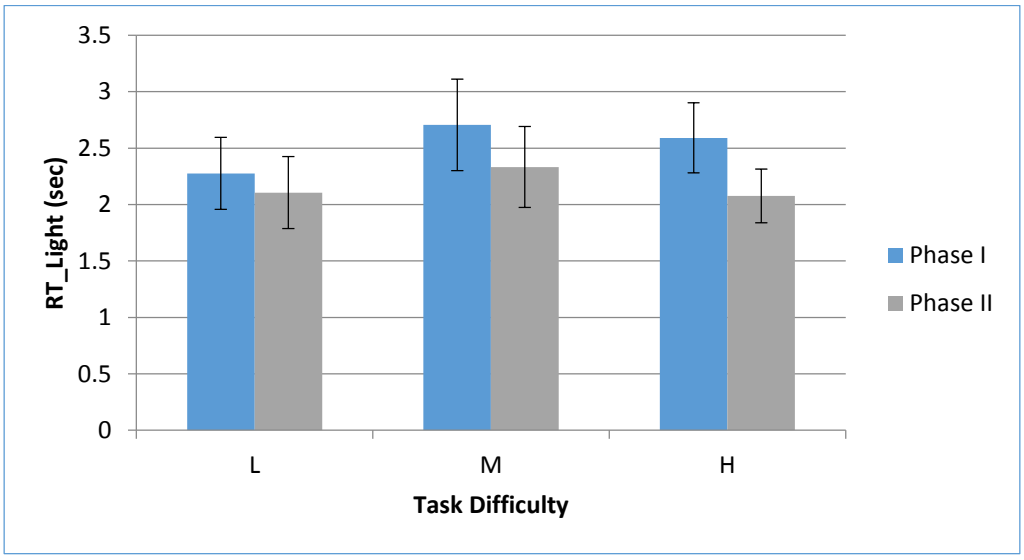
7 Results and Discussion

7.1 Comparison between Two Phases

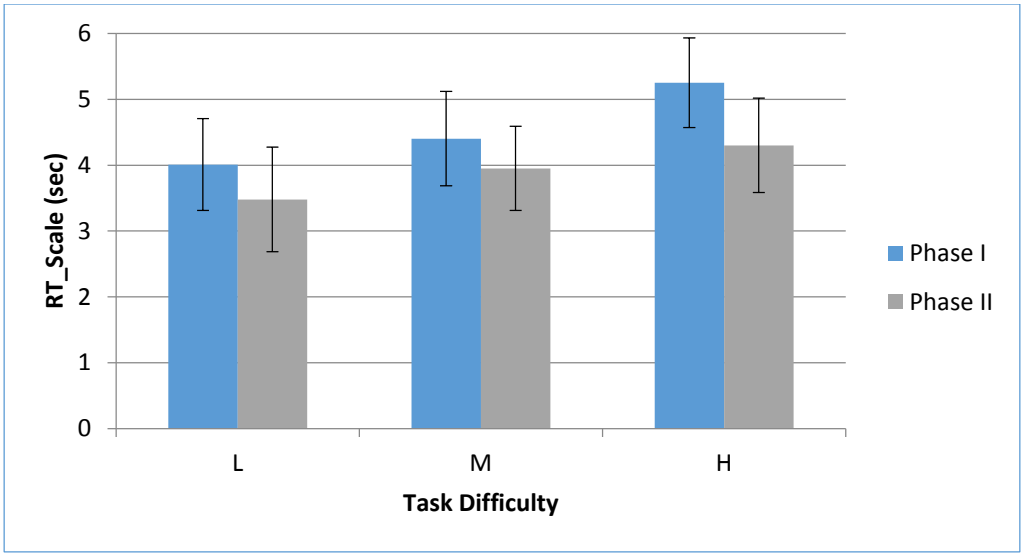
After the rearrangement of task weights in Phase II, user performance was recorded and compared with the baseline performance obtained from Phase I (see Table 5.4 and 5.5). As mentioned in section 6.2, paired comparisons of multitasking performance were conducted to validate the improvement of user performance after the rearrangement of task weights. The descriptive statistics of user performance are summarized in Appendix F. Each response variable included all three levels of task difficulty. The error bars of all bar charts in this section represent the standard deviation of corresponding response mean.

7.1.1 RT of MATB-II Tasks

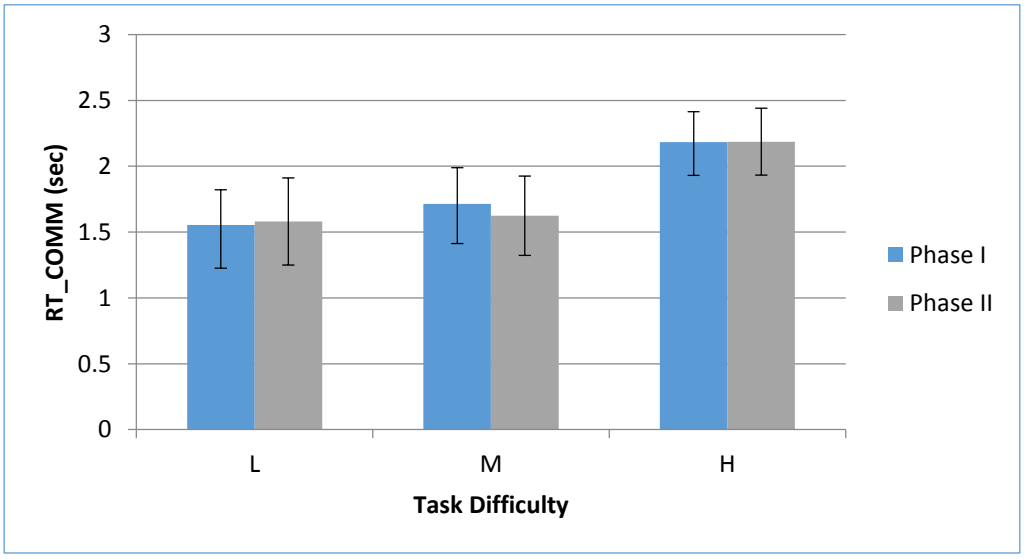
The bar charts in Figure 7.1 present the comparisons of average RT of MATB-II tasks between Phase I and Phase II, including Light, Scale, and COMM tasks (RT_Light, RT_Scale, and RT_COMM).



(a) Comparison of RT_Light



(b) Comparison of RT_Scale



(c) Comparison of RT_COMM

Figure 7.1 Comparisons of RT of MATB-II tasks between two phases

Note: L = Low; M = Medium; H = High.

From the bar charts in Figure 7.1, a trend of decrease in RT was observed across all three levels of task difficulty for all three tasks. To further analyze the improvement between two phases, paired t-tests with Bonferroni correction (Rice, 2006) were conducted. Table 7.1 shows a summary of results of the paired t-tests for RT of MATB-II tasks between Phase I and Phase II ($p < 0.05/3 = 0.0167$, three task difficulty levels).

Table 7.1 Summary of Paired t-test Results for RT

	Task Difficulty	N	DF	t-Ratio	p-value
RT_Light	**Low**	50	49	-2.36	0.0111*
	Medium	50	49	-4.84	<0.0001**
	High	50	49	-10.13	<0.0001**
RT_Scale	**Low**	50	49	-2.22	0.0155*
	Medium	50	49	-2.37	0.0109*
	High	50	49	-6.57	<0.0001**
RT_COMM	**Low**	50	49	0.25	0.6
	Medium	50	49	-0.96	0.1716
	High	50	49	0.04	0.5139

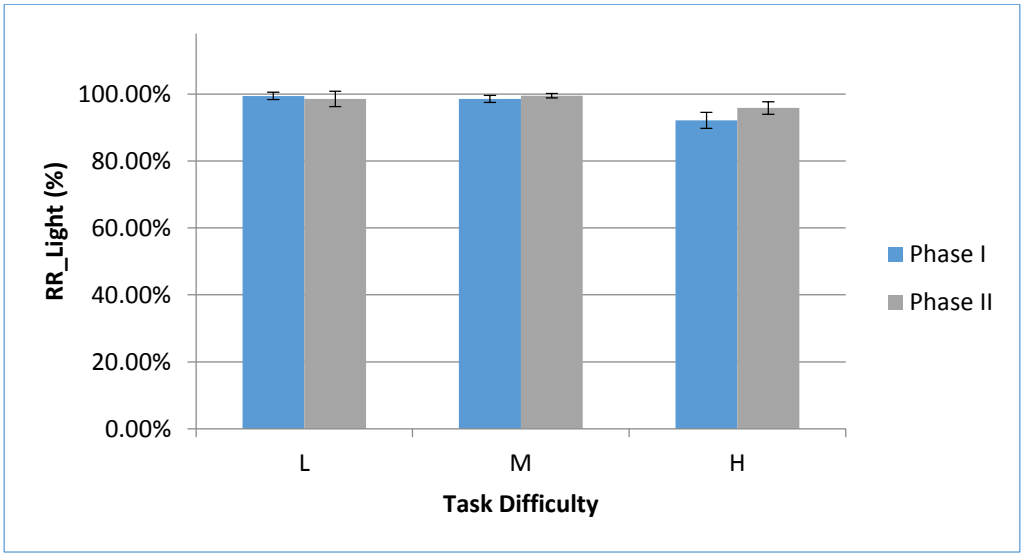
Note. * significance level: $p < 0.0167$. ** significance level: $p < 0.0001$.

The results in Figure 7.1 and Table 7.1 implied the improvement of task performance by showing the decrease of RT in Phase II. As stated in the first hypothesis (H1(a)), RTs of MATB-II tasks would be lower in Phase II than in Phase I. From the results of paired t-tests, RTs of Light and Scale tasks showed significant decrease ($p < 0.0167$) in second phase across all levels of task difficulty which supported hypothesis H1(a). For COMM task, significant improvement was not found for RT, but the results showed that the RT maintained at a stable range for each level of task

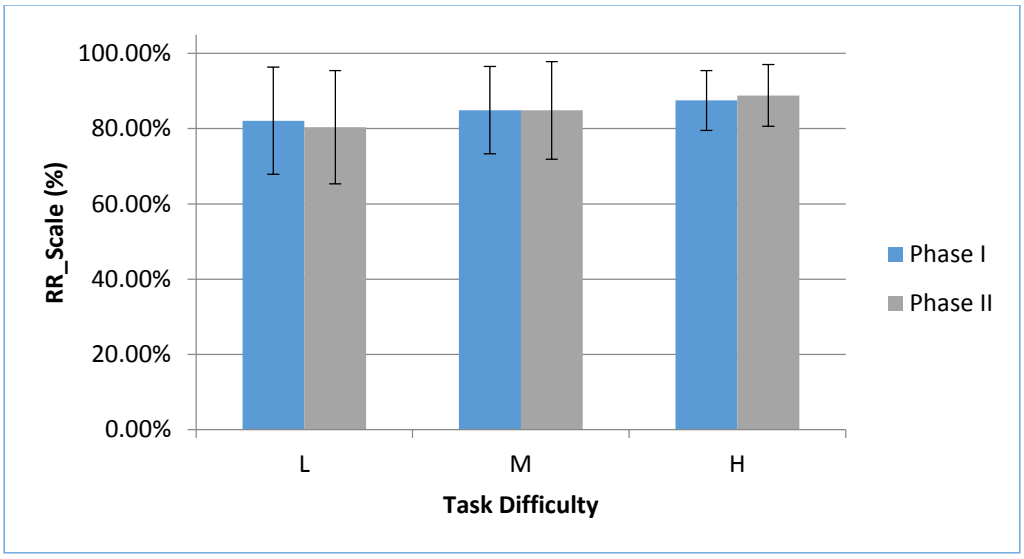
difficulty (e.g., Low: 1.55s vs. 1.58s; Medium: 1.71s vs. 1.62s; High: 2.18s vs. 2.19s, for Phase I vs. Phase II, respectively).

7.1.2 RR of MATB-II Tasks

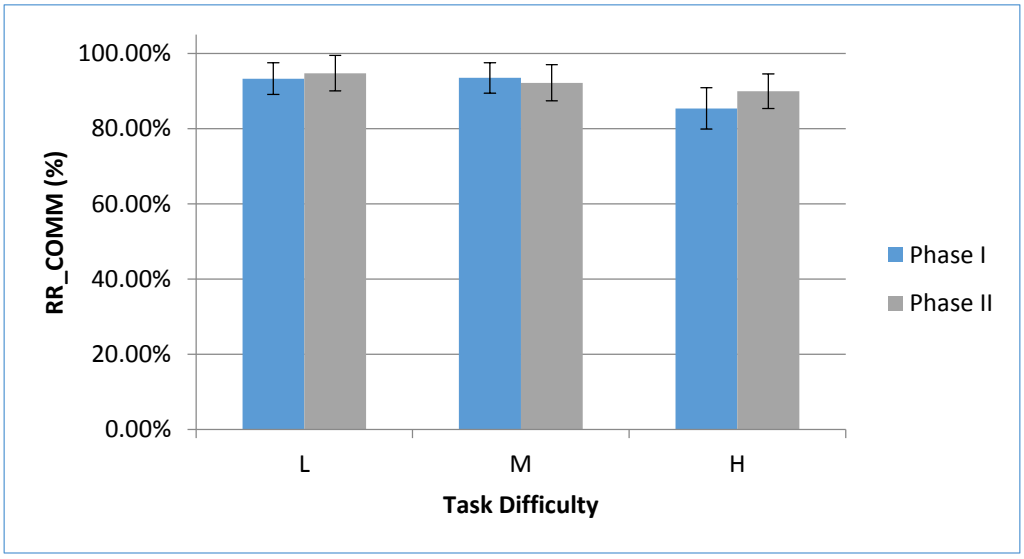
The bar charts in Figure 7.2 present the comparisons of average RR of MATB-II tasks between Phase I and Phase II, including Light, Scale, COMM, and RMAN tasks (RR_Light, RR_Scale, RR_COMM, and RR_RMAN).



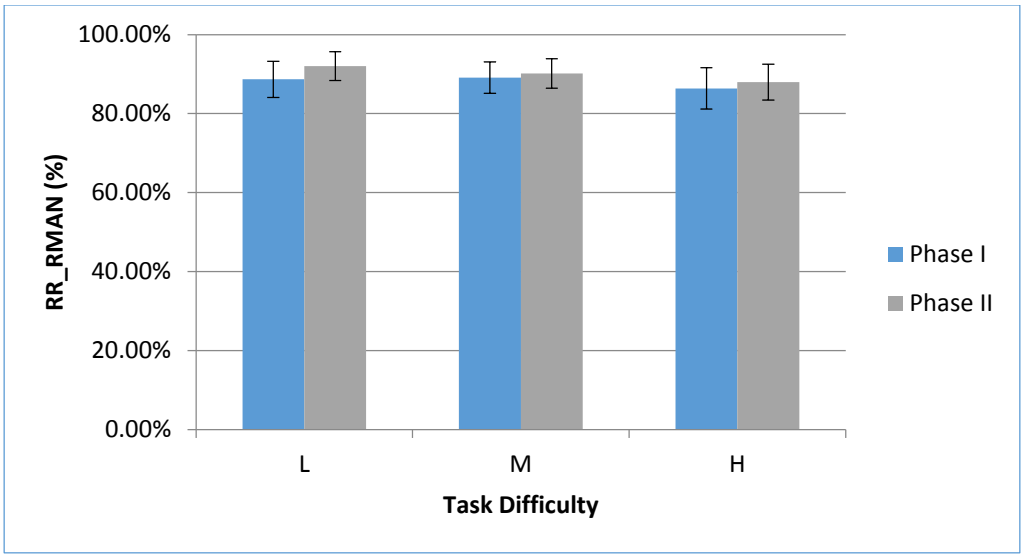
(a) Comparison of RR_Light



(b) Comparison of RR_Scale



(c) Comparison of RR_COMM



(d) Comparison of RR_RMAN

Figure 7.2 Comparisons of RR of MATB-II tasks between two phases

Note: L = Low; M = Medium; H = High.

From the bar charts in Figure 7.2, no general trend of RR was observed for MATB-II tasks. The descriptive statistics (see Appendix F) implied that RR for MATB-II tasks maintained at a small range between two phases across all three levels of task difficulty. To further investigate if there was an improvement of RR between two phases, paired t-tests with Bonferroni correction (Rice, 2006) were conducted. Table 7.2 shows a summary of results of the paired t-tests for RR of MATB-II tasks between Phase I and Phase II ($p < 0.05/3 = 0.0167$, three task difficulty levels).

Table 7.2 Summary of Paired t-test Results for RR

	Task Difficulty	N	DF	t-Ratio	p-value
RR_Light	**Low**	50	49	-1.13	0.1331
	Medium	50	49	3.03	0.0019*
	High	50	49	5.50	<0.0001*
RR_Scale	**Low**	50	49	-0.38	0.3543
	Medium	50	49	-0.01	0.4953
	High	50	49	0.62	0.2708
RR_COMM	**Low**	50	49	0.81	0.201
	Medium	50	49	-0.77	0.7781
	High	50	49	2.36	0.0223
RR_RMAN	**Low**	50	49	2.66	0.0052*
	Medium	50	49	0.90	0.1861
	High	50	49	1.25	0.1087

Note. * significance level: p < 0. 0167. ** significance level: p < 0.0001.

From the results in Table 7.2, only few significant results were detected for RR in MATB-II. None of the descriptive statistics or paired t-tests showed a general direction of RR between two phases. One of the assumptions of the proposed model was that the response ratio (RR) for MATB-II tasks would keep at a constant value for each subtask (see section 4.3), which meant that RR for each MATB-II subtask would maintain at the same level between two phases. After the practice session, participants who reach a stable level of RR for MATB-II tasks were invited to following sessions of this study. During experiment sessions in both Phase I and II, participants maintained a relative high level of overall RR (> 80%) for all MATB-II tasks. From the descriptive statistics (Table 5.4 and Table 6.1) and the comparisons between two phases (Figure 7.1 and Table 7.1), a small range of the overall RR was maintained between Phase I and Phase II across all three levels of task difficulty for each subtask (e.g., RRs of MATB-II tasks between Phase I and Phase

II: Light: 92% – 100%; Scale: 82% – 89%; COMM: 85% – 95%; RMAN: 86% – 92%). The results of RR of MATB-II tasks in two phases validated the assumption of the approach of task weight rearrangement in the proposed model.

7.1.3 RMSD of TRACK Task

The bar charts in Figure 7.3 present the comparisons of mean RMSD of TRACK tasks between Phase I and Phase II.

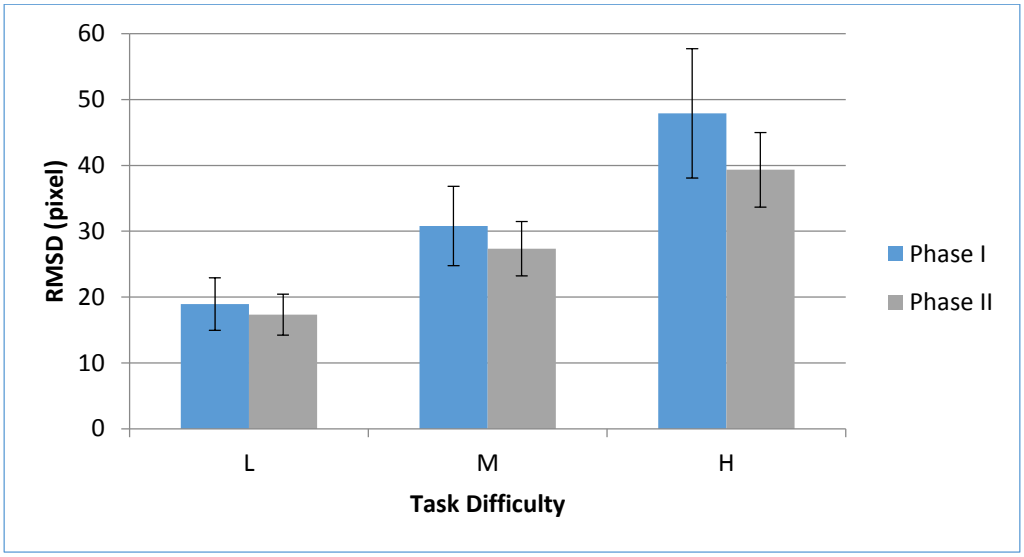


Figure 7.3 Comparisons of RMSD between Phase I and II

Note: L = Low; M = Medium; H = High.

From Figure 7.3, a trend of decrease of RMSD was implied for all three levels of task difficulty between two phases. To further analyze the differences in RMSD between two phases, paired t-tests with Bonferroni correction (Rice, 2006) were conducted. Table 7.3 shows a summary

of results of the paired t-tests for RMSD between Phase I and Phase II ($p < 0.05/3 = 0.0167$, three task difficulty levels).

Table 7.3 Summary of Paired t-test Results for RMSD

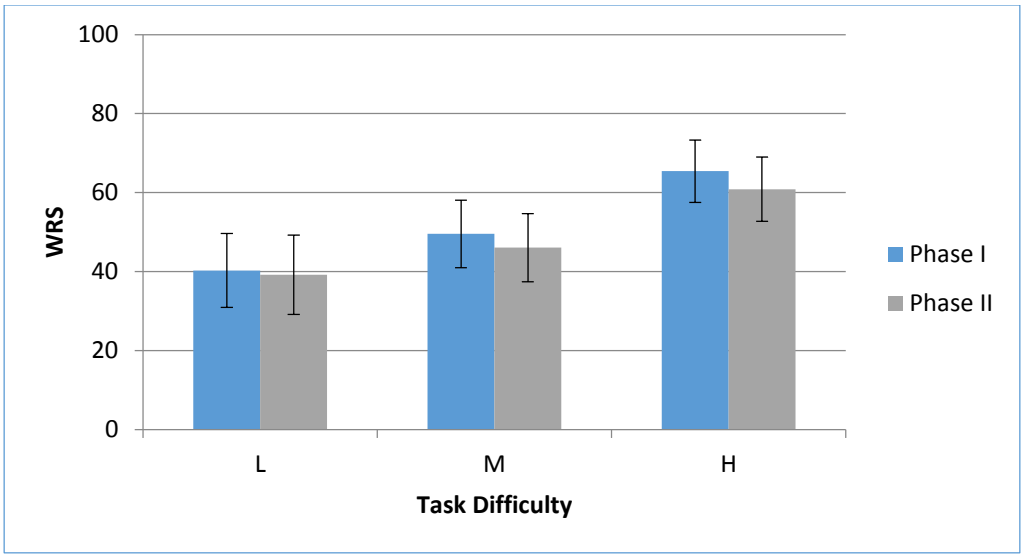
	Task Difficulty	N	DF	t-Ratio	p-value
RMSD	**Low**	50	49	-2.17	0.0173
	Medium	50	49	-3.95	0.001*
	High	50	49	-5.57	<0.0001**

Note. * significance level: $p < 0.0167$. ** significance level: $p < 0.0001$.

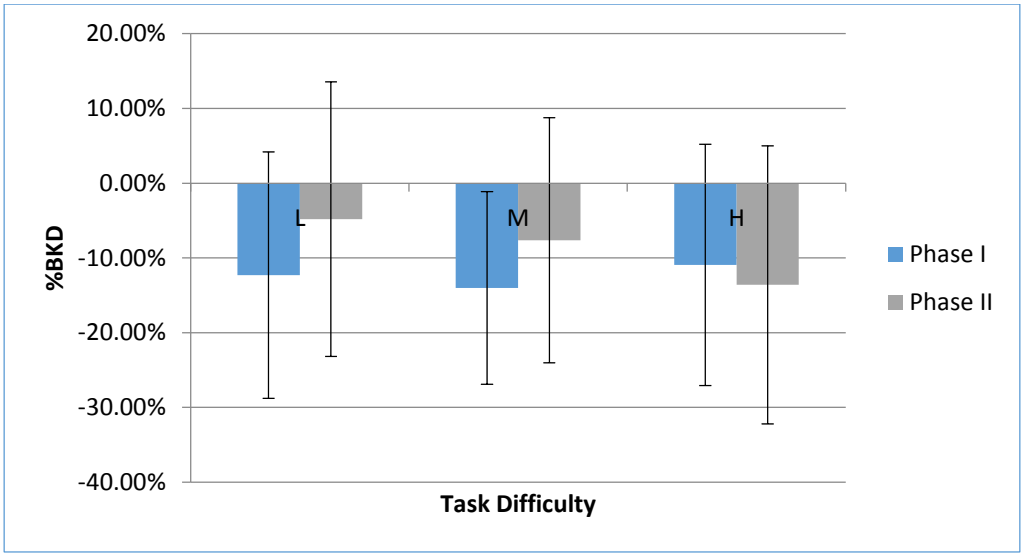
As stated in hypothesis H1(b), RMSD of TRACK task would be lower in Phase II than in Phase I. The results in Table 7.3 showed a significant or marginally significant decrease in RMSD at each level of task difficulty. This result indicated the improvement of user performance on the tracking task during Phase II.

7.1.4 Subjective Workload Rating and Eye Measures

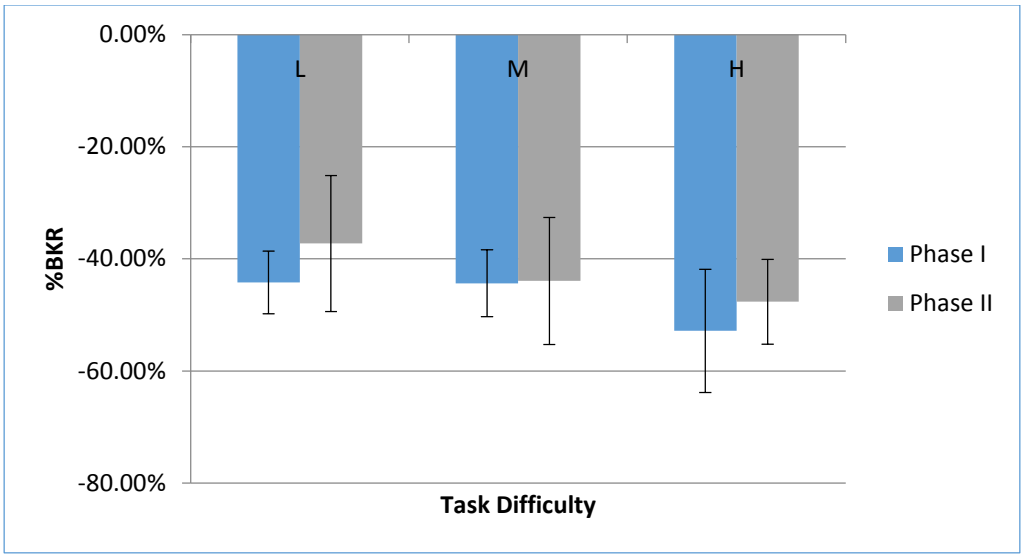
The descriptive statistics of subjective workload rating (WRS) and eye measures (%BKR and %BKD) were summarized in Appendix F. The bar charts in Figure 7.4 present the comparisons of workload between Phase I and Phase II, including WRS, %BKR and %BKD.



(a) Comparison of WRS



(b) Comparison of %BKD



(c) Comparison of %BKR

Figure 7.4 Comparisons of Workload between Phase I and II

From the bar charts in Figure 7.4, both WRS and %BKR showed a decreasing trend from Phase I to Phase II among all three levels of task difficulty. Paired t-tests with Bonferroni correction (Rice, 2006) were conducted to further investigate the changes in workload between two phases. Table 7.4 shows a summary of results of the paired t-tests for workload between Phase I and Phase II ($p < 0.05/3 = 0.0167$, three task difficulty levels).

Table 7.4 Summary of Paired t-test Results for Workload

	Task Difficulty	N	DF	t-Ratio	p-value
WRS	**Low**	50	49	-0.593	0.2779
	Medium	50	49	-1.624	0.0554
	High	50	49	-2.6313	0.0057*
%BKD	**Low**	50	49	0.98	0.1661
	Medium	50	49	1.37	0.177
	High	50	49	1.08	0.1423
%BKR	**Low**	50	49	0.01	0.4963
	Medium	50	49	0.79	0.2177
	High	50	49	1.49	0.071

Note. * significance level: $p < 0.0167$. ** significance level: $p < 0.0001$.

Hypothesis H1(c) stated that the mental workload of the operators would be lower in Phase II than in Phase I. From the comparison of WRS between two phases, a significant decrease was found for WRS at high level ($p = 0.0057$) of task difficulty. WRS did not show a significant decrease between two phases at the low task difficulty level (e.g., Low: 40.29 (Phase I) vs. 39.20 (Phase II)) or medium level (e.g., Medium: 49.54 (Phase I) vs. 46.07 (Phase II)), which is likely due to the ease of completing the task at these difficulty levels.

The average blink rate and blink duration showed a decrease from baseline trial during the experiment. Their standard deviations were relatively large and no statistical trend can be concluded here. From paired t-tests of eye measures, no significant results were found for %BKD or %BKR.

7.2 Effects of Task Difficulty and WMC

The effects of task difficulty and working memory capacity (WMC) on multitasking performance for both Phase I and Phase II of the experiment were investigated by performing ANOVAs. The response variables included MATB-II task performance, subjective workload rating, and eye measures. In the statistical model for this split-plot design, WMC was the whole plot factor and task difficulty was the split plot factor. Since task difficulty included three levels, Tukey's HSD tests were performed if significant effects of task difficulty were found for any response variables.

7.2.1 RT of MATB-II Tasks

During the two-phase experiment, RT were collected for Light, Scale, and COMM tasks. Table 7.5 shows the significant effects of task difficulty and WMC on RT in MATB-II tasks.

Table 7.5 Summary of ANOVA Results for RT

	Effect	F-value	p-value
RT_Light	Task Difficulty	F(2, 295) = 5.93	0.003**
	WMC	F(1, 295) = 7.62	0.0061*
RT_Scale	Task Difficulty	F(2, 295) = 12.98	<0.0001**
	WMC	F(1, 295) = 5.60	0.0186*
RT_COMM	Task Difficulty	F(2, 295) = 38.21	<0.0001**
	WMC	F(1, 295) = 20.90	<0.0001**

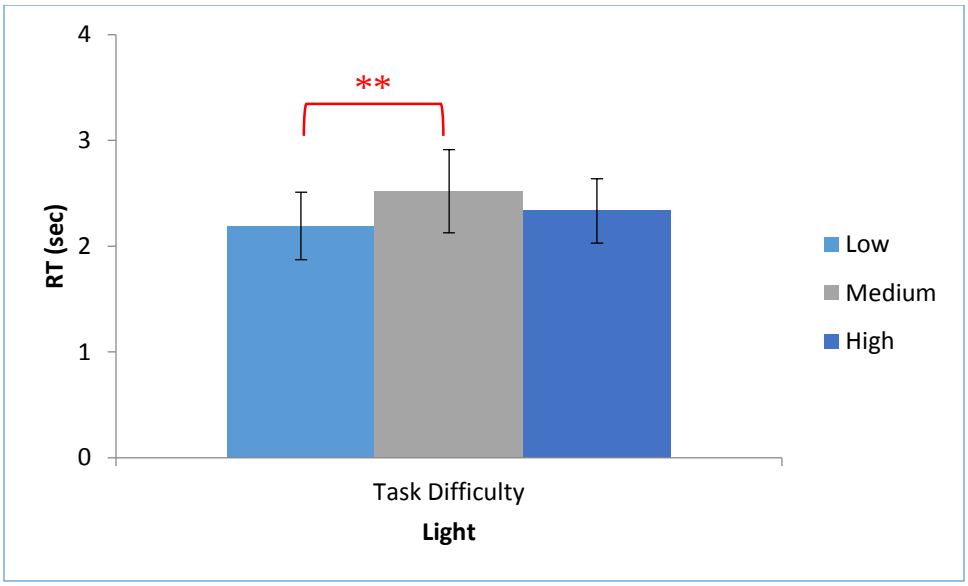
Note. * significance level: $p < 0.05$. ** significance level: $p < 0.0001$.

The results from ANOVA revealed significant effects for both task difficulty and WMC on RT of MATB-II tasks. No significant interaction effect was found. The whole plot main effect,

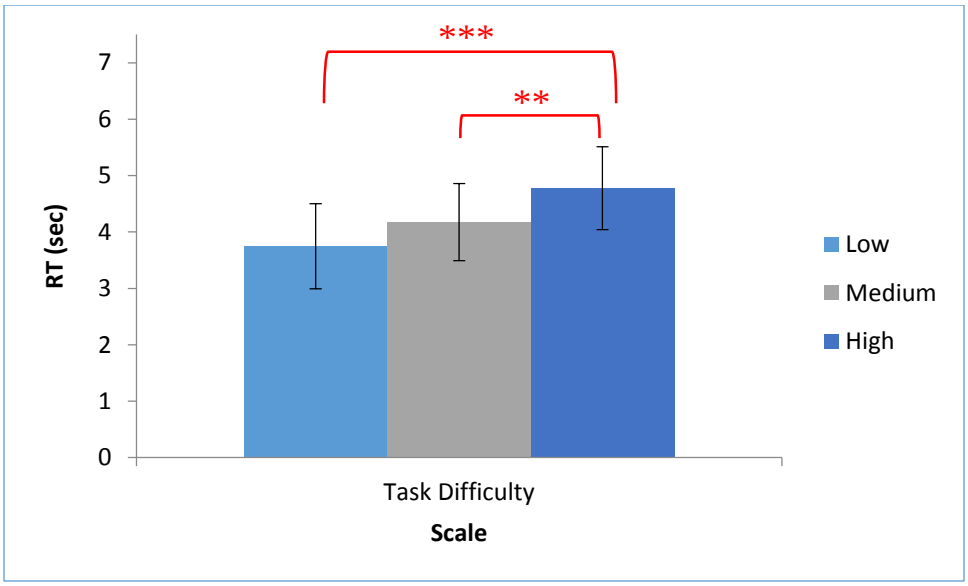
WMC, showed significant effects on RT in three MATB-II tasks, Light ($F(1, 295) = 7.62$, $p = 0.0061$), Scale ($F(1, 295) = 5.60$, $p = 0.0186$), and COMM ($F(1, 295) = 20.90$, $p < 0.0001$). These results showed that participants with high WMC had a lower RT of MATB-II tasks than those with low WMC which supported hypothesis H3(b).

The split plot main effect, task difficulty, also showed significant effects on RT in three MATB-II tasks, Light ($F(2, 295) = 5.93$, $p = 0.003$), Scale ($F(2, 295) = 12.98$, $p < 0.0001$), and COMM ($F(2, 295) = 38.21$, $p < 0.0001$). These findings implied the increase of RT with the increase of task difficulty for MATB-II tasks which supported hypothesis H2(b).

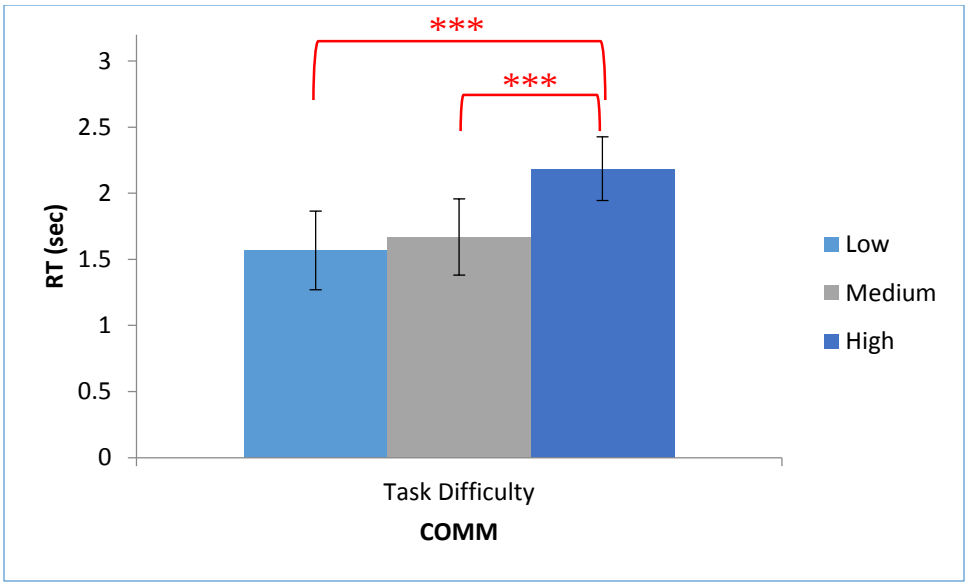
Tukey's HSD test was performed for RT of each task and results are illustrated in Figure 7.5. For Light task, RT at low level (mean = 2.19, SD = 0.64) was significantly shorter than RT at medium level (mean = 2.52, SD = 0.78). For Scale task, a trend of increase of RT was detected and RT at high level (mean = 4.78, SD = 1.47) was significantly longer than RTs at low level (mean = 3.74, SD = 1.51) and medium level (mean = 4.18, SD = 1.37). For COMM task, a trend of increase of RT was also revealed and RT at high level (mean = 2.16, SD = 0.48) was significantly longer than RTs at low level (mean = 1.57, SD = 0.60) and medium level (mean = 1.67, SD = 0.58).



(a) Main effect of task difficulty on RT of Light task



(b) Main effect of task difficulty on RT of Scale task



(c) Main effect of task difficulty on RT of COMM task

Figure 7.5 Results of Tukey's HSD Test for RT

Note: ** $p < 0.05$; *** $p < 0.0001$

7.2.2 RR of MATB-II Tasks

During the two-phase experiment, RR were collected for Light, Scale, COMM, and RMAN tasks. Table 7.6 shows the significant effects of task difficulty and WMC on RR in MATB-II tasks.

Table 7.6 Summary of ANOVA Results for RR

	Effect	F-value	p-value
RR_Light	Task Difficulty	F(2, 295) = 68.57	<0.0001**
	WMC	F(1, 295) = 1.52	0.2187
	Task Difficulty *WMC	F(2, 295) = 5.05	0.007**
RR_Scale	Task Difficulty	F(2, 295) = 2.12	0.1224
	WMC	F(1, 295) = 2.35	0.1262
RR_COMM	Task Difficulty	F(2, 295) = 13.24	<0.0001**
	WMC	F(1, 295) = 8.10	0.0047*
RR_RMAN	Task Difficulty	F(2, 295) = 3.92	0.0208*
	WMC	F(1, 295) = 19.97	<0.0001**

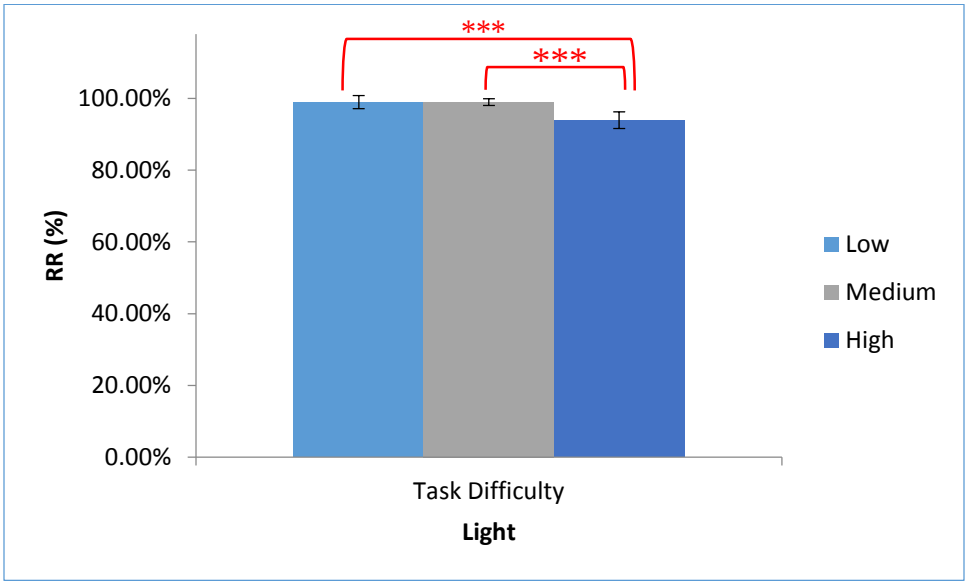
Note. * significance level: $p < 0.05$. ** significance level: $p < 0.0001$.

The results from ANOVA indicated several significant effects of task difficulty and WMC on RR of MATB-II tasks. One significant interaction effect, task difficulty and WMC, was found for RR of Light task ($F(2, 295) = 5.05$, $p = 0.007$).

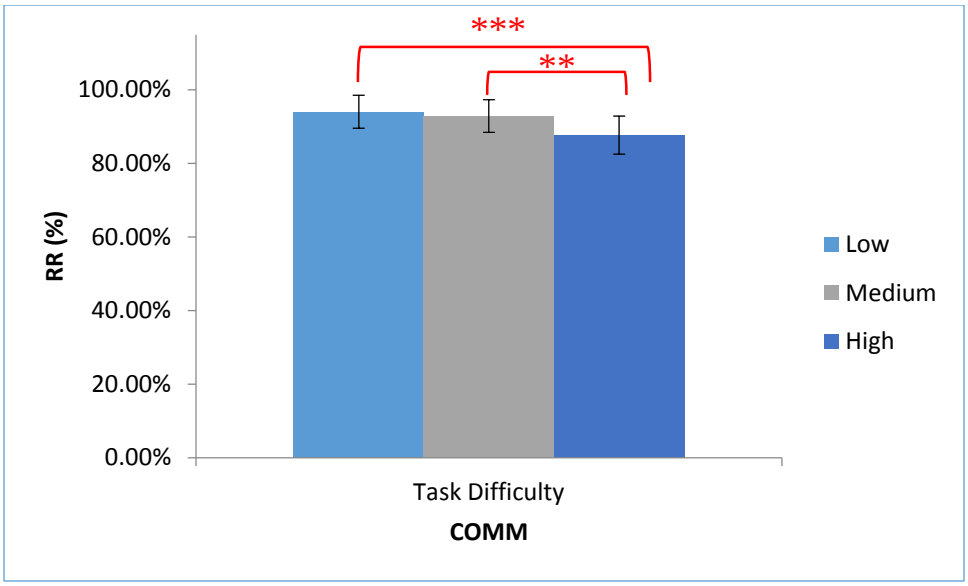
The whole plot main effect, WMC, showed significant effects on RR in two MATB-II tasks, COMM ($F(2, 295) = 8.10$, $p = 0.0047$) and RMAN ($F(2, 295) = 19.97$, $p < 0.0001$). These results indicated that participants with high WMC produced a higher RR of MATB-II tasks than those with low WMC which supported part of hypothesis H3(a).

The split plot main effect, task difficulty, showed significant effects on RR in three MATB-II tasks, Light ($F(2, 295) = 68.57$, $p < 0.0001$), COMM ($F(2, 295) = 13.24$, $p < 0.0001$), and RMAN ($F(2, 295) = 3.92$, $p = 0.0208$). These findings implied the decrease of RR with the increase of task difficulty for MATB-II tasks which supported hypothesis H2(a).

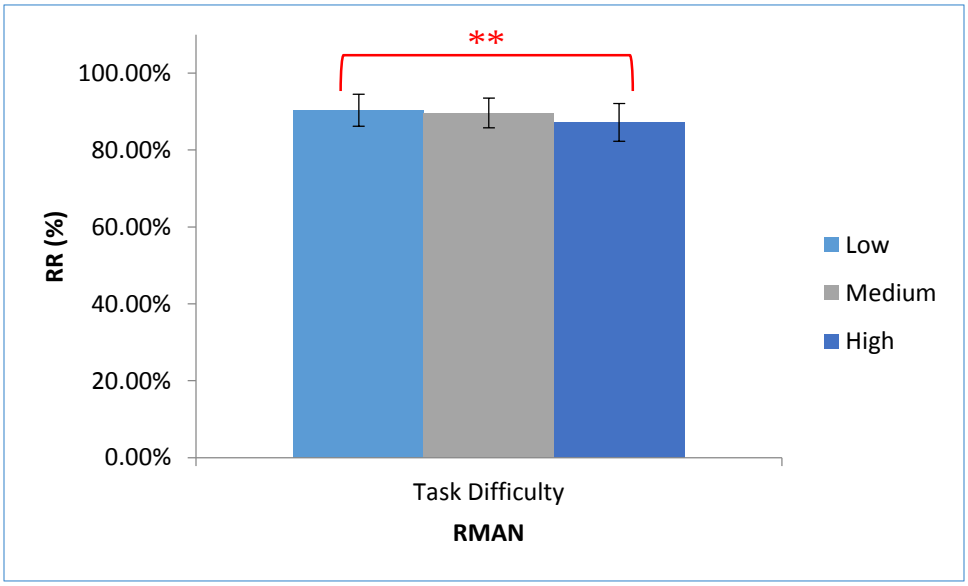
Tukey's HSD test was performed for RR of Light, COMM, and RMAN tasks. The results are illustrated in Figure 7.6. For Light task, RR at high level (mean = 93.97%, SD = 4.65%) was significantly lower than RR at low level (mean = 98.99%, SD = 3.62%) and medium level (mean = 99.03%, SD = 1.85%). For COMM task, RR at high level (mean = 87.67%, SD = 10.31%) was also significantly lower than RR at low level (mean = 94.01%, SD = 8.94%) and medium level (mean = 92.84%, SD = 8.86%). For RMAN task, RR at high level (mean = 87.18%, SD = 9.79%) was also significantly lower than RR at low level (mean = 90.34%, SD = 8.39%).



(a) Main effect of task difficulty on RR of Light task



(b) Main effect of task difficulty on RR of COMM task



(c) Main effect of task difficulty on RR of RMAN task

Figure 7.6 Results of Tukey's HSD Test for RR

Note: ** $p < 0.05$; *** $p < 0.0001$

7.2.3 RMSD of TRACK Task

Performance of TRACK task was assessed by the value of RMSD during the experiment. Table 7.7 shows the ANOVA results for RMSD.

Table 7.7 Summary of ANOVA Results for RMSD

	Effect	F-value	p-value
RMSD	Task Difficulty	F(2, 295) = 113.71	<0.0001**
	WMC	F(1, 295) = 1.58	0.2104

Note. * significance level: $p < 0.05$. ** significance level: $p < 0.0001$.

A significant effect of the split plot main effect, task difficulty, was indicated from the results of ANOVA. No significant interaction effect was found. During TRACK task, RMSD was significantly increased by the rise in task difficulty ($F(2, 295) = 113.71$, $p < 0.0001$). This result supported hypothesis H2(c).

Tukey's HSD test was performed for RMSD of TRACK task and results are illustrated in Figure 7.7. Significant differences in RMSD were found among all three levels of task difficulty. The value of RMSD at high level (mean = 43.61, SD = 16.51) was significantly higher than RMSD at low level (mean = 18.13, SD = 7.15) and medium level (mean = 29.07, SD = 10.41) in TRACK task.

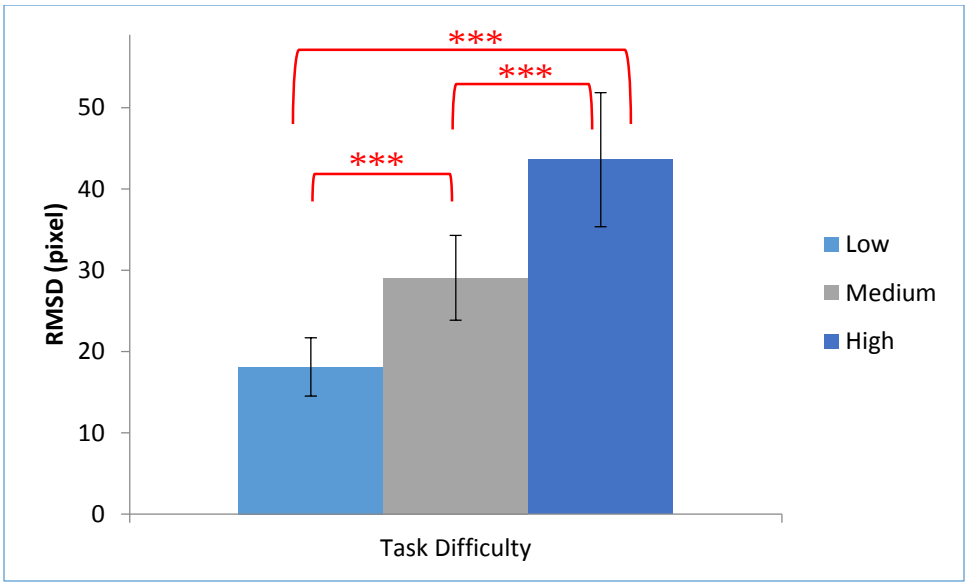


Figure 7.7 Results of Tukey's HSD Test for RMSD

Note: ** $p < 0.05$; *** $p < 0.0001$

7.2.4 Subjective Workload Rating and Eye Measures

During the experiment, subjective workload rating scale (WRS) and eye measures (blink rate and blink duration) were collected to estimate mental workload of participants. Table 7.8 shows the significant effects of task difficulty and WMC on these measures in MATB-II tasks.

Table 7.8 Summary of ANOVA Results for Workload

	Effect	F-value	p-value
WRS	Task Difficulty	F(2, 295) = 45.90	<0.0001**
	WMC	F(1, 295) = 3.08	0.0805
%BKD	Task Difficulty	F(2, 295) = 1.00	0.3676
	WMC	F(1, 295) = 4.60	0.0327*
%BKR	Task Difficulty	F(2, 295) = 2.12	0.1215
	WMC	F(1, 295) = 1.86	0.174

Note. * significance level: $p < 0.05$. ** significance level: $p < 0.0001$.

The whole plot main effect, WMC, showed a significant effect on %BKD ($F(2, 295) = 4.60$, $p = 0.0327$). This result indicated that participants with higher WMC had a longer blink duration during the experiment than participants with lower WMC which supported part of hypothesis H3(d). No significant effect of WMC on WRS and %BKR was found.

A significant effect of the split plot main effect, task difficulty, was found for WRS from the results of ANOVA. No significant interaction effect was found. During the experiment, participants' WRS was significantly increased with the increase of task difficulty ($F(2, 295) = 45.90$, $p < 0.0001$). This result supported hypothesis H2(d). No significant effect of task difficulty on blink duration (%BKD) and blink rate (%BKR) was found.

Tukey's HSD test was performed for WRS and results are illustrated in Figure 7.8. Significant differences in WRS were found among all three levels of task difficulty. The WRS score at high level (mean = 63.15, SD = 16.13) was significantly higher than WRS score at low level (mean = 39.74, SD = 19.34) and medium level (mean = 47.80, SD = 17.20).

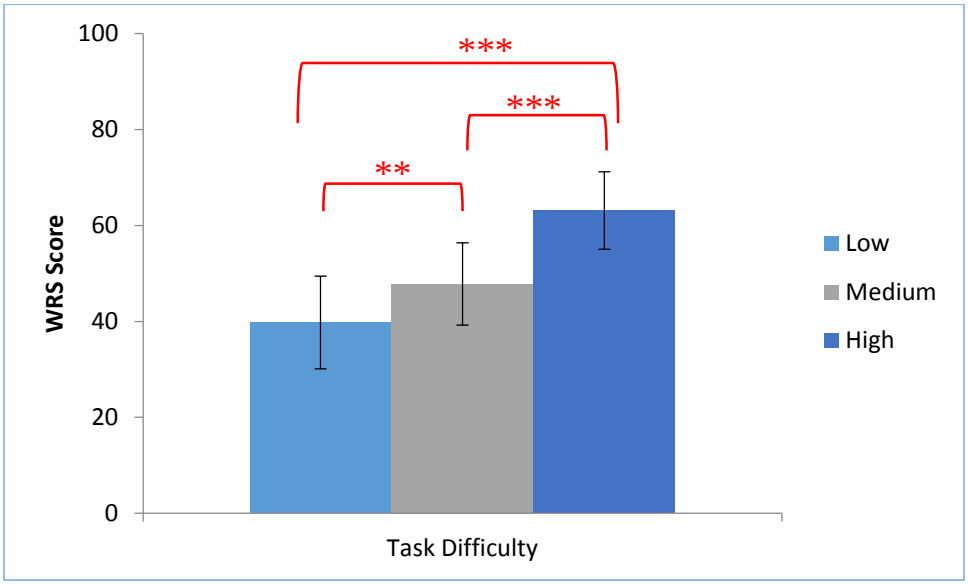


Figure 7.8 Results of Tukey's HSD Test for WRS

Note: ** $p < 0.05$; *** $p < 0.0001$

7.3 Correlation Analysis

To investigate the cognitive relations among the subtasks in the multitasking platform, MATB-II, correlation analysis was performed for all MATB-II task performance. Since response variables of task performance were not normally distributed, a nonparametric correlation analysis, Spearman's correlation analysis, was applied. The response variables include RT for Light, Scale, and COMM tasks (RT_Light, RT_Scale, and RT_COMM), RR for Light, Scale, COMM, and RMAN tasks (RR_Light, RR_Scale, RR_COMM, and RR_RMAN), and RMSD of TRACK task. Table 7.9 shows the correlations among task performance in MATB-II ordered by Spearman coefficient.

Table 7.9 Correlations among MATB-II Task Performance

Response Variables		Spearman's ρ	p-value
RT_Scale	RT_Light	0.5461	<.0001**
RMSD	RT_COMM	0.5146	<.0001**
RR_COMM	RR_Light	0.3281	<.0001**
RMSD	RT_Scale	0.3227	<.0001**
RR_Scale	RR_Light	0.2638	<.0001**
RR_RMAN	RR_Light	0.192	0.0008*
RR_RMAN	RR_COMM	0.1567	0.0065*
RMSD	RT_Light	0.1313	0.0229*
RT_COMM	RT_Scale	0.1161	0.0446*
RR_COMM	RR_Scale	0.019	0.7432
RT_COMM	RT_Light	0.0056	0.9232
RR_COMM	RT_Light	-0.0373	0.52
RT_COMM	RR_Scale	-0.0374	0.5192
RR_RMAN	RT_Scale	-0.0686	0.2362
RR_COMM	RT_Scale	-0.0889	0.1246
RR_RMAN	RMSD	-0.0964	0.0957
RR_RMAN	RR_Scale	-0.0971	0.0933
RR_RMAN	RT_Light	-0.1081	0.0615
RR_Light	RT_Light	-0.1822	0.0015*
RR_RMAN	RT_COMM	-0.2124	0.0002*
RMSD	RR_Scale	-0.2163	0.0002*
RR_COMM	RT_COMM	-0.3205	<.0001**
RR_Scale	RT_Light	-0.3457	<.0001**
RMSD	RR_COMM	-0.363	<.0001**
RR_Light	RT_Scale	-0.3689	<.0001**
RR_Scale	RT_Scale	-0.369	<.0001**
RT_COMM	RR_Light	-0.3936	<.0001**
RMSD	RR_Light	-0.5477	<.0001**

Note. * significance level: $p < 0.05$. ** significance level: $p < 0.0001$.

The multitasking platform, MATB-II, provides a set of tasks which have cognitive relations. The correlation coefficients in Table 7.9 shows that most pairs of task performance were

highly correlated. For instance, RMSD of TRACK task was highly correlated with RR of Light, Scale, and COMM tasks, and RT of Light, Scale, COMM tasks. RT between Light and Scale tasks, Scale and COMM tasks were highly correlated. RR between Light and COMM tasks, Light and Scale tasks, Light and RMAN tasks, COMM and RMAN tasks were highly correlated. These results validated the strong cognitive relations among the tasks in MATB-II.

7.4 General Discussion

7.4.1 Model improvement

The first research goal of this study was to propose and validate a model to quantify the information transmitted and processed between human operator and system, and provide the approach to evaluate and improve human performance in a multitasking environment. A quantitative model that converted the stimuli from system and human operator into a baud rate (bps) was proposed in this study based information theory (Shannon, 1948). An approach, task rearrangement, to improve individuals' user performance in multitasking was included in the proposed model based on the cross-entropy optimization (Fang, Rajasekera, & Tsao, 1997). To validate the proposed model, a two-phase experiment was conducted and hypotheses were formulated in terms of the improvement of user performance after task rearrangement in the experiment. A practice session was conducted for each participant to eliminate the carryover effect of training in multitasking.

Statistical comparison between the two phases revealed that most of the response variables verify the performance improvement in Phase II across all three levels of task difficulty. After performed task rearrangement based on the performance baseline obtained from Phase I, a shorter

RT and a lower RMSD of MATB-II tasks were detected among all three levels of task difficulty in Phase II experiment. For the RR of MATB-II tasks, all participants reached a stable relative high RR after practice session and maintained their RR at a small range during the two-phase experiment. This outcome validated one of the assumption of the proposed model that RR would maintain at a constant value between two phases.

With regard to the changes in mental workload, significant decrease of subjective workload rating was found at high level of task difficulty in Phase II and a trend of decreased workload at low and medium levels was also implied by the descriptive statistics. Blink rate showed a trend of increase among all three levels of task difficulty in Phase II, but blink duration failed to show a constant change between two phases. Similar as suggested in previous study (Back et al., 1994), blink rate and blink duration may be selected as the indexes of the presence of workload, but not efficient for discriminating among levels of mental demand.

7.4.2 User performance and workload

The second research goal of this study was to investigate the relationship between WMC, task difficulty, and human performance in a multitasking environment. During the two-phase experiment, task difficulty was manipulated by the overall baud rate of the system and WMC was assessed by the automated OSPAN test (Unsworth et al., 2005). Two hypotheses were formulated to investigate their relationship.

Task difficulty was seen to be an effective method to manipulate demand in multitasking, which was observed not only through task performance metrics, but also through subjective workload rating during the experiment. From the statistical analyses, the second hypothesis (H2)

was supported that task difficulty would affect user performance in multitasking. With the increase of task difficulty, task performance showed significant decrease (e.g., RT increased, RR decreased, and RMSD increased), except for RR of Scale task. This may due to the ease of Scale task and the RR kept a constant high level during the experiment. In addition, subjective rating demonstrated an increase of workload with the increase of task difficulty, however, eye measures failed to show similar results.

The individual differences in WMC has been examined as an important predictor of varying cognitive abilities (Kane & Engle, 2003; Engle & Kane, 2004). This statement has been verified by the results from the experiment in this study. Individual differences in WMC were estimated by the automated OSPAN test and the participant pool spanned an ample range of OSPAN scores (28 – 75 on a scale of 0 – 75). The third hypothesis (H3) surmised that different working memory capacities (WMC) would yield different levels of multitasking performance. Various user performance metrics were developed by different participants and verified the assumption in H3. For example, participants with high WMC produced higher level of MATB-II task performance (e.g., lower RT, higher RR, and lower RMSD) and lower level of subjective workload than participants whit low WMC. However, significant results were limited in terms of the effect of WMC on user performance in this study. This may due to a limited population pool (25 participants from college students) comparing to other studies (Goldinger et al., 2003; Kane et al., 2001; Kane & Engle, 2003; Watson et al., 2005).

8 Conclusion

8.1 Summary of Findings

The objectives of this study were to propose and validate a quantitative model for user performance and improvement in a multitasking environment; and to investigate the relationship between working memory capacity (WMC), task difficulty, and multitasking performance. To accomplish these research goals, a quantitative model based on information theory was proposed, consisting of the quantification of stimuli from multitasking system as baud rate (bps), selection of task difficulty and task weight, as well as the rearrangement of task weights with respect to a user-specific performance baseline. This study applied a two-phase experimental approach. Phase I applied the proposed model and identified a performance baseline for each individual in a multitasking environment, MATB-II. Phase II validated the proposed model by using a rearranged set of multitasks for each individual. The effects of WMC and task difficulty on multitasking performance had been analyzed with experimental data from both of these phases. Correlation analysis was applied and implied the cognitive relations among each task in MATB-II.

Three hypotheses were established based on the research goals. Twenty-five participants met the experiment criteria and completed the two phases experiment. The results from the experiment verified the hypotheses. By following the approach of task weight rearrangement in the proposed model, an improvement of task performance and a decrease of mental workload were reported in Phase II, as compared to the baseline performance obtained from Phase I. This finding supported hypothesis H1 and revealed that the proposed model can effectively quantify system stimuli and user's operation. In addition, it showed a practical approach to enhance individual performance in multitasking.

Results from this study also demonstrated the relationship between WMC, task difficulty, and multitasking performance. Previous research (Adler & Benbunan-Fich, 2015; Wickens, 1991; Rupp et al., 2014) showed that task difficulty was an effective method to manipulate mental demand in multitasking. In this study, task difficulty was manipulated by different levels of overall baud rate based on the proposed quantitative model. Results from experiment implied that user task performance decreased and their subjective workload increased with the increase of task difficulty. These results verified hypothesis H2. Individual differences in WMC were evaluated by the automated OSPAN test in this study. Results from experiment verified the effect of WMC on task performance and subjective workload in a multitasking environment. These results supported most parts of hypothesis H3. These findings can provide insight into personnel selection for some highly-complex and safety-critical professions.

8.2 Limitations of Present Study

Several aspects of the experiment and model assumptions may limit the implement of the proposed quantitative model in other task environments or platforms. First, the baud rate for each MATB-II task was set up as a constant value for each trial and the time interval between signals kept consistent across the experiment trial. This method represented a simplified approach for the set-up of task baud rate. However, in practice, the time interval between signals may fluctuate throughout the operation and generate a random value of baud rate from time to time. In this case, the baud rate of task will fail to remain a constant value and adjustment of the baud rate set-up in the proposed model would be desired.

In the experiment, task difficulty was manipulated by the overall baud rate and included three difficulty levels. Although the results showed significant effects of task difficulty on the majority of task performance metrics, the structure of levels of task difficulty for each subtask was relatively simple. For each level of task difficulty, all subtasks were set up at the same level to obtain individual performance baseline in Phase I. Different combinations of task difficulty for subtasks should be took into account for a complex multitasking environment.

A parallel strategy in multitasking was applied during the experiment. Participants were instructed to complete each subtask as quickly and as accurately as possible by switching between MATB-II subtasks. They were allowed to leave a subtask incomplete and switch to another subtask during their operations. However, current data analyses are unable to show that if the participants achieved true parallel performance which seemed challenging due to limited human attention (Salvucci & Taatgen, 2011).

During this study, a single multitasking system, MATB-II, was applied. The significant correlations among the MATB-II subtask performance measures make clear the dependence of the task outcomes and confirm the single-system nature of the simulation. Consequently, the modeling approach developed in this study should be considered as having utility for single-system performance. Care should be taken in interpreting the results of modeling approach for any multiple-system scenarios.

Additionally, this research only recruited twenty-five participants from campus. This participant population may limit the results of individual differences in multitasking since only college students were included in the experiment. To acquire a wide range of WMC for the study of individual differences, an extensive population (> 100) is recommended (Goldinger et al., 2003; Kane et al., 2001; Kane & Engle, 2003; Watson et al., 2005).

8.3 Research Applications and Future Work

The main research application of this study is to provide an approach for the quantification of information and to improve user performance in a multitasking environment. A quantitative model including a task rearrangement approach was introduced in this study and a two-phase experiment demonstrated the approach to imply the proposed model in a multitasking platform. Improvement of user performance was observed from second phase of experiment. These results verified the assumptions of the proposed model and validated the feasibility of proposed approach. The assumptions applied in this study were defined by the requirement of MATB-II operation. With minor adjustment based on different system requirements, the proposed model can be applied in other multitasking environments to quantify and improve user performance.

Taking the limitations into account, to assess the information content in a system more accurately, future study should pursue an approach to estimate the baud rate for a complex system with a dynamic status. The multitasking system should include different combinations of tasks and the instant baud rate of each task may keep fluctuating throughout the operation.

During the experiment, participants were instructed to perform a parallel strategy that same task priority was assigned for all subtasks and they switch among tasks. The effectiveness of this strategy has not been examined yet in this study. Future work should verify if task strategy is truly achieved by the human operator. For example, the sequential pattern mining can be applied to tack participants' gaze pattern and determine the sequences of task operation (Ayres, Flannick, Gehrke, & Yiu, 2002; Steichen, Wu, Toker et al., 2014) to validate the strategy.

This research also provided insight into personnel selection in terms of the individual differences in WMC. To obtain a better estimate of WMC of target population, future work should

include a wide-ranging population to assess the WMC and set up criteria for personnel selection. It should be interesting to investigate other aspects of individual differences in multitasking using the proposed quantitative model. Future study may evaluate other characteristics of individual differences and investigate the performance improvement by applying the proposed optimization approach.

By following the proposed model, improvement of user performance was demonstrated in this study. As Box's (1976) maxim "All models are wrong but some are useful", further improvements in the proposed model have potential to continue to improve user performance in a multitasking environment.

REFERENCES

Adler, R. F., & Benbunan-Fich, R. (2015). The effects of task difficulty and multitasking on performance. *Interacting with Computers*, 27(4), 430-439.

Alexopoulou, P., Morris, A., & Hepworth, M. (2014). A New Integrated Model for Multitasking during Web Searching. *Procedia - Social and Behavioral Sciences*, *147*, 16–25.

Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (2001). Intelligenz-Struktur-Test 2000 R. *Göttingen: Hogrefe*, 2.

Anderson, J. B., & Johnnesson, R. (2006). *Understanding information transmission* (Vol. 18). John Wiley & Sons.

Aricò, P., Borghini, G., Graziani, I., Taya, F., Sun, Y., Bezerianos, A., ... & Babiloni, F. (2014, August). Towards a multimodal bioelectrical framework for the online mental workload evaluation. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE* (pp. 3001-3004). IEEE.

Ayres, J., Flannick, J., Gehrke, J., & Yiu, T. (2002, July). Sequential pattern mining using a bitmap representation. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 429-435). ACM.

Backs, R. W., Ryan, A. M., & Wilson, G. F. (1994). Psychophysiological measures of workload during continuous manual performance. *Human Factors*, 36(3), 514-531.

Backs, R. W., & Seljos, K. A. (1994). Metabolic and cardiorespiratory measures of mental effort : the effects of level of difficulty in a working memory task, *16*, 57–68.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *The Psychology of Learning and Motivation Advances in Research and Theory*.

Bahrick, H. P., Noble, M., & Fitts, P. M. (1954). Extra-task performance as a measure of learning a primary task. *Journal of Experimental Psychology*, 48(4), 298.

Bailey, B. P., & Iqbal, S. T. (2008). Understanding Changes in Mental Workload during Execution of Goal-Directed Tasks and Its Application for Interruption Management, *14*(4).

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(4), 693-709.

Baldwin, C. L., & Penaranda, B. N. (2012). Adaptive training using an artificial neural network and EEG metrics for within- and cross-task workload classification. *NeuroImage*, *59*(1), 48–56.

Basak, C., Boot, W. R., Voss, M. W., & Kramer, A. F. (2008). Can Training in a Real-Time Strategy Videogame Attenuate cognitive decline in older adults? *Psychology and Aging*, *23*(4), 765–777.

Benbunan-Fich, R., Adler, R. F., & Mavlanova, T. (2011). Measuring multitasking behavior with activity-based metrics. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 18(2), 7.

Benedetto, S., Pedrotti, M., Minin, L., Baccino, T., Re, A., & Montanari, R. (2011). Driver workload and eye blink duration. *Transportation Research Part F: Traffic Psychology and Behaviour*, *14*(3), 199–208.

Bhavsar, P., Srinivasan, B., & Srinivasan, R. (2015). Pupillometry Based Real-Time Monitoring of Operator's Cognitive Workload to Prevent Human Error during Abnormal Situations. *Industrial and Engineering Chemistry Research*, *55*(12), 3372–3382.

Bharwani, G. (2006). *The application of quantitative information processing to modeling and simulation of industrial workplace tasks.* Doctoral dissertation.

Bluedorn, A. C., Kaufman, C. F., & Lane, P. M. (1992). How many things do you like to do at once? An introduction to monochronic and polychronic time. *The Executive*, 6(4), 17-26.

Boiten, F., Sergeant, J., & Geuze, R. (1992). Event-related desynchronization: the effects of energetic and computational demands. *Electroencephalography and clinical neurophysiology*, 82(4), 302-309.

Borghini, G., Aricò, P., Astolfi, L., Toppi, J., Cincotti, F., Mattia, D., … Babiloni, F. (2013). Frontal EEG theta changes assess the training improvements of novices in flight simulation tasks. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, (1), 6619–6622.

Borghini, G., Aricò, P., Graziani, I., Salinari, S., Sun, Y., Taya, F., … Babiloni, F. (2015). Quantitative Assessment of the Training Improvement in a Motor-Cognitive Task by Using EEG, ECG and EOG Signals. *Brain Topography*.

Bowers, M. A., Christensen, J. C., & Eggemeier, F. T. (2014). The Effects of Workload Transitions in a Multitasking Environment. *Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting2*, 220–224.

Box, G. E. (1976). Science and statistics. *Journal of the American Statistical Association*, 71(356), 791-799.

Brenner, M., Doherty, E. T., & Shipp, T. (1994). Speech measures indicating workload demand. *Aviation, space, and environmental medicine.*

Brookings, J. B. (1990). A Confirmatory Factor Analytic Study of Time-Sharing Performance and Cognitive Abilities, *59*, 43–59.

Brouwer, A. M., Hogervorst, M. A., Van Erp, J. B., Heffelaar, T., Zimmerman, P. H., & Oostenveld, R. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of neural engineering*, 9(4), 045008.

Brumby, D. P., Howes, A., & Salvucci, D. D. (2007, April). A cognitive constraint model of dual-task trade-offs in a highly dynamic driving task. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 233-242). ACM.

Brumby, D. P., Salvucci, D. D., & Howes, A. (2009). Focus on driving. *Proceedings of the 27th International Conference on Human Factors in Computing Systems - CHI 09*, (January), 1629.

Bühner, M., König, C. J., Pick, M., & Krumm, S. (2006). Working memory dimensions as differential predictors of the speed and error aspect of multitasking performance. *Human Performance*, 19(3), 253-275.

Camden, A., Nickels, M., Fendley, M., & Phillips, C. A. (2016). A case for information theory-based modelling of human multitasking performance. *Theoretical Issues in Ergonomics Science*, 18(3), 266-278.

Card, S. K., English, W. K., & Burr, B. J. (1978). Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. *Ergonomics*, 21(8), 601-613.

Card, S. K., Newell, A., & Moran, T. P. (1983). The psychology of human-computer interaction.

Chan, R. B., & Childress, D. S. (1990). On information transmission in human-machine systems: channel capacity and optimal filtering. *IEEE transactions on systems, man, and cybernetics*, 20(5), 1136-1145.

Clark, R. E., & Estes, F. (1996). Cognitive task analysis for training. *International Journal of Educational Research*, 25(5), 403-417.

Clark, D., & Ivry, R. B. (2010). Multiple systems for motor skill learning. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(4), 461–467.

Clegg, B. A., Vieane, A. Z., Wickens, C. D., Gutzwiller, R. S., & Sebok, A. L. (2014, September). The effects of automation-induced complacency on fault diagnosis and management performance in process control. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, No. 1, pp. 844-848). Sage CA: Los Angeles, CA: SAGE Publications.

Colom, R., Martínez-Molina, A., Shih, P. C., & Santacreu, J. (2010). Intelligence, working memory, and multitasking performance. *Intelligence*, *38*(6), 543–551.

Comstock, J. R., & Arnegard, R. J. (1992). Multi-Attribute Task Battery for Human Operator Workload and Strategic Behavior Research. *NASA Technical Memorandum*, (January).

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic bulletin & review*, 12(5), 769-786.

Corbett, B., Yamaguchi, T., Liu, S., Huang, L., Bahn, S., & Nam, C. S. (2013, July). Influence of haptic feedback on a pointing task in a haptically enhanced 3D virtual environment. In *International Conference on Human-Computer Interaction* (pp. 561-567). Springer Berlin Heidelberg.

Crossman, E. R. F. W. (1960). The information-capacity of the human motor-system in pursuit tracking. *Quarterly Journal of Experimental Psychology*, 12(1), 01-16.

Damos, D. (1986). The effect of using voice generation and recognition systems on the performance of dual tasks. *Ergonomics*, 29(11), 1359-1370.

Damos, D. (1991). Dual-task methodology: Some common problems. *Multiple-task performance*, 101-119.

David, P., Xu, L., Srivastava, J., & Kim, J. H. (2013). Media multitasking between two conversational tasks. *Computers in Human Behavior*, *29*(4), 1657–1663.

De Rivecourt, M., Kuperus, M. N., Post, W. J., & Mulder, L. J. (2008). Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight. *Ergonomics*, *51*(9), 1295–1319.

De Waard, D. (1996). *The measurement of drivers' mental workload*. Netherlands: Groningen University, Traffic Research Center.

Dehais, F., Causse, M., & Pastor, J. (2008, March). Embedded eye tracker in a real aircraft: new perspectives on pilot/aircraft interaction monitoring. In *Proceedings from the 3rd International Conference on Research in Air Transportation*. Fairfax, USA: Federal Aviation Administration.

Delbridge, K. A. (2000). *Individual differences in multi-tasking ability: Exploring a nomological network* (Doctoral dissertation, Michigan State University. Dept. of Psychology).

Di Stasi, L. L., Antolí, A., & Cañas, J. J. (2013). Evaluating mental workload while interacting with computer-generated artificial environments. *Entertainment Computing*, 4(1), 63-69.

Du, J. T., & Spink, A. (2011). Toward a Web Search Model : Integrating Multitasking, Cognitive Coordination , and Cognitive Shifts, *62*(1977), 1446–1472.

Durkee, K., Geyer, A., Pappada, S., Ortiz, A., & Galster, S. (2013). Real-time workload assessment as a foundation for human performance augmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8027 LNAI*, 279–288.

Dux, P. E., Tombu, M. N., Harrison, S., Rogers, B. P., Tong, F., & Marois, R. (2009). Training Improves Multitasking Performance by Increasing the Speed of Information Processing in Human Prefrontal Cortex. *Neuron*, *63*(1), 127–138.

Dzubak, C. M. (2008). Multitasking: The good, the bad, and the unknown. *The Journal of the Association for the Tutoring Profession*, 1(2), 1-12.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: a test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(5), 972.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *Psychology of learning and motivation*, 44, 145-199.

Epps, B. W. (1986, September). Comparison of six cursor control devices based on Fitts' law models. In *Proceedings of the Human Factors Society Annual Meeting* (Vol. 30, No. 4, pp. 327-331). Sage CA: Los Angeles, CA: SAGE Publications.

Erickson, K. I., Colcombe, S. J., Wadhwa, R., Bherer, L., Peterson, M. S., Scalf, P. E., … Kramer, A. F. (2007). Training-induced functional activation changes in dual-task processing: An fMRI study. *Cerebral Cortex*, *17*(1), 192–204.

Fang, S. C., Rajasekera, J. R., & Tsao, H. S. J. (1997). Entropy optimization and mathematical programming. *International Series in Operations Research & Management Science, 8*.

Fischer, R., & Plessow, F. (2015). Efficient multitasking: parallel versus serial processing of multiple tasks. *Frontiers in psychology*, 6.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, 47(6), 381.

Fogarty, C., & Stern, J. A. (1989). Eye movements and blinks: their relationship to higher cognitive processes. *International Journal of Psychophysiology*, 8(1), 35-42.

Folds, D. J., Gerth, J. M., & Engelman, W. R. (1987). *Enhancement of human performance in manual target acquisition and tracking*. Georgia Institute of Tech Atlanta Systems Engineering Lab.

Fournier, L. R., Wilson, G. F., & Swain, C. R. (1999). Electrophysiological, behavioral, and subjective indexes of workload when performing multiple tasks: Manipulations of task difficulty and training. *International Journal of Psychophysiology*, *31*(2), 129–145.

Garson, G. D. (2012). Testing statistical assumptions. *Asheboro, NC: Statistical Associates Publishing.*

Gevins, A., & Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science*, 4(1-2), 113-131.

Goldinger, S. D., Kleider, H. M., Azuma, T., & Beike, D. R. (2003). "Blaming the victim" under memory load. *Psychological Science*, *14*(1), 81–85. http://doi.org/10.1111/1467-9280.01423

Guastello, A. D., Guastello, S. J., & Guastello, D. D. (2013). Personality trait theory and multitasking performance: implications for ergonomic design. *Theoretical Issues in Ergonomics Science*, (May 2015), 1–19.

Gundel, A., & Wilson, G. F. (1992). Topographical changes in the ongoing EEG related to the difficulty of mental tasks. *Brain topography*, 5(1), 17-25.

Gutzwiller, R. S., Wickens, C. D., & Clegg, B. a. (2014). Workload overload modeling: An experiment with MATB-II to inform a computational model of task management. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 849–853.

Haider, H., & Frensch, P. A. (2002). Why aggregated learning follows the power law of practice when individual learning does not: Comment on Rickard (1997, 1999), Delaney et al. (1998), and Palmeri (1999).

Hambrick, D. Z., Oswald, F. L., Darowski, E. S., Rench, T. A., & Brou, R. (2010). Predictors of multitasking performance in a synthetic work paradigm. *Applied Cognitive Psychology*, *24*(8), 1149–1167.

Hankins, T. C., & Wilson, G. F. (1998). A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight. *Aviation, space, and environmental medicine*, 69(4), 360-367.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, 52, 139-183.

Haykin, S. S. (1988). *Digital communications* (pp. 445-446). New York: Wiley.

Hertzum, M., & Holmegaard, K. D. (2013). Perceived Time as a Measure of Mental Workload: Effects of Time Constraints and Task Success. *International Journal of Human-Computer Interaction*, *29*(1), 26–39.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, 4(1), 11-26.

Higgins, J. J. (2003). Introduction to modern nonparametric statistics.

Horrey, W. J., & Wickens, C. D. (2004). Driving and side task performance: The effects of display clutter, separation, and modality. *Human factors*, 46(4), 611-624.

Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of applied psychology*, 85(6), 869-879.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, *45*(3), 188–196.

Ikuma, L. H., Harvey, C., Taylor, C. F., & Handal, C. (2014). A guide for assessing control room operator performance using speed and accuracy, perceived workload, situation awareness, and eye tracking. *Journal of Loss Prevention in the Process Industries*, *32*, 454–465.

Imants, P., & de Greef, T. (2014, September). Eye Metrics for Task-Dependent Automation. In *Proceedings of the 2014 European Conference on Cognitive Ergonomics* (p. 23). ACM.

Jagacinski, R. J., & Monk, D. L. (1985). Fitts' Law in Two dimensions with hand and head movements movements. *Journal of Motor Behavior*, 17(1), 77-95.

Johnson, E. J., Bellman, S., & Lohse, G. L. (2003). Cognitive lock-in and the power law of practice. *Journal of Marketing*, 67(2), 62-75.

Jorna, P. G. (1992). Spectral analysis of heart rate and psychological state: A review of its validity as a workload index. *Biological psychology*, 34(2), 237-257.

Junco, R., & Cotten, S. R. (2012). No A 4 U: The relationship between multitasking and academic performance. *Computers and Education*, *59*(2), 505–514.

Kaber, D. B., Kim, S. H., & Wang, X. (2011). Computational Cognitive Modeling of Human-Robot Interaction Using a GOMS Methodology, 53–75.

Kaber, D. B., & Kim, S. H. (2011). Understanding Cognitive Strategy with Adaptive Automation in Dual-Task Performance Using Computational Cognitive Models. *Journal of Cognitive Engineering and Decision Making*, *5*(3), 309–331.

Kaber, D. B., Wang, X., & Kim, S.-H. (2006). Computational Cognitive Modeling of Operator Behavior in Telerover Navigation. *2006 IEEE International Conference on Systems, Man and Cybernetics*, 3210–3215.

Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, 130(2), 169.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*(1), 47–70.

Kantowitz, B. H., & Elvers, G. C. (1988). Fitts' law with an isometric controller: effects of order of control and controldisplay gain. *Journal of Motor Behavior*, 20(1), 53-66.

Ke, Y., Qi, H., He, F., Liu, S., Zhao, X., Zhou, P., … Ming, D. (2014). An EEG-based mental workload estimator trained on working memory task can work well under simulated multi-attribute task. *Frontiers in Human Neuroscience*, *8*(September), 703.

Kidwell, B. D., Miller Jr, W. D., & Parasuraman, R. (2014, September). Automation Complacency: Using Non-Invasive Brain Stimulation to Change Attention Allocation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, No. 1, pp. 240-244). Sage CA: Los Angeles, CA: SAGE Publications.

Kieras, D. E., & Meyer, D. E. (1995). Predicting human performance in dual-task tracking and decision making with computational models using the EPIC architecture. In *Proceedings of the First International Symposium on Command and Control Research and Technology*, National Defense University, June. Washington, DC: National Defense University.

Kim, J. W., Ritter, F. E., & Koubek, R. J. (2013). An integrated theory for improved skill acquisition and retention in the three stages of learning. *Theoretical Issues in Ergonomics Science*, *14*(1), 22–37.

Klein, K., & Fiss, W. H. (1999). The reliability and stability of the Turner and Engle working memory task. *Behavior Research Methods, Instruments, & Computers*, 31(3), 429-432.

König, C. J., Buhner, M., & Murling, G. (2005). Working Memory, Fluid Intelligence, and Attention Are Predictors of Multitasking Performance, but Polychronicity and Extraversion Are Not. *Human Performance*, *18*(3), 243–266.

Kruchinin, A. Y. (2015). Multitasking system for real time pattern recognition. *Automation and Remote Control*, *76*(1), 166–171.

La Pointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(6), 1118.

Leidheiser, W., & Pak, R. (2014, September). The Effects of Age and Working Memory Demands on Automation-Induced Complacency. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, No. 1, pp. 1919-1923). SAGE Publications.

Li, H., Wickens, C. D., Sarter, N., & Sebok, A. (2014). Stages and levels of automation in support of space teleoperations. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 56(6), 1050-1061.

Lin, L. (2013). Multiple Dimensions of Multitasking Phenomenon. *International Journal of Technology and Human …*, *9.1*(January-March 201), 37–49.

Lin, T., & Imamiya, A. (2006). Evaluating usability based on multimodal information. *Proceedings of the 8th International Conference on Multimodal Interfaces - ICMI '06*, 364.

Lin, Y., Zhang, W. J., & Watson, L. G. (2003). Using eye movement parameters for evaluating human-machine interface frameworks under normal control operation and fault detection situations. *International Journal of Human Computer Studies*, *59*(6), 837–873.

Liu, S., Wadeson, A., Kim, N. Y., & Nam, C. S. (2016, September). Effects of Working Memory Capacity, Task Switching, and Task Difficulty on Multitasking Performance. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 60, No. 1, pp. 502-506). SAGE Publications.

Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychological review*, 108(2), 393.

MacKenzie, I. S. (1992). Fitts' law as a research and design tool in human-computer interaction. *Human-computer interaction*, 7(1), 91-139.

Maclin, E. L., Mathewson, K. E., Low, K. A., Boot, W. R., Kramer, A. F., Fabiani, M., & Gratton, G. (2011). *Learning to multitask: effects of video game practice on electrophysiological indices of attention and resource allocation.* Psychophysiology, 48(9), 1173-1183.

Mateo, J. C., Cowgill Jr, J. L., Moore Jr, T. J., Gilkey Jr, R. H., Simpson Jr, B. D., Weisenberger Jr, J. M., & Manning Jr, J. T. (2005, September). Evaluation of a collaborative movement task in a distributed three-dimensional virtual environment. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 49, No. 17, pp. 1578-1582). Sage CA: Los Angeles, CA: SAGE Publications.

Miller, S. (2001). Workload measures. *The University of Iowa*, (August), 1–65.

Miller, W. D. (2010). The U.S. Air Force-Developed Adaptation of the Multi-Attribute Task Battery for the Assessment of Human Operator Workload and Strategic.

Miller, W. D., Schmidt, K. D., Estepp, J. R., & Bowers, M. (2014). An Updated Version of the U.S. Air Force Multi-Attribute Task Battery (AF-MATB).

Monsell, S. (2003). Task switching. *Trends in cognitive sciences*, 7(3), 134-140.

Montgomery, D. C. (1997). *Montgomery Design and Analysis of Experiments*. John Wiley.

Murata, A., & Iwase, H. (2001). Extending Fitts' law to a three-dimensional pointing task. *Human movement science*, 20(6), 791-805.

Murray, K. B., & Bellman, S. (2011). Productive play time: The effect of practice on consumer demand for hedonic experiences. *Journal of the Academy of Marketing Science*, *39*(3), 376–391.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. *Cognitive Skills and Their Acquisition*, *6*, 1–55.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive psychology*, 7(1), 44-64.

Oberauer, K., Süß, H. M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity—facets of a cognitive ability construct. *Personality and Individual Differences*, 29(6), 1017-1045.

Oberauer, K., Süß, H. M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence*, 31(2), 167-193.

Oksama, L., & Hyönä, J. (2016). Position tracking and identity tracking are separate systems: Evidence from eye movements. *Cognition*, 146, 393-409.

Oldfield, R. C. (1971). The assessment and analysis of handedness : Edinburgh inventory, *9*.

Onnasch, L., Ruff, S., & Manzey, D. (2012, September). Operators' Adaption to Unreliability of Alarm Systems: A Performance and Eye-Tracking Analysis. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 56, No. 1, pp. 248-252). Sage Publications.

Oswald, F. L., Hambrick, D. Z., & Jones, L. A. (2007). Keeping all the plates spinning: Understanding and predicting multitasking performance. *Learning to solve complex scientific problems*, 77-97.

Palmen, T. J. (1999). Theories of automaticity and the power law of practice.

Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(2), 324.

Phillips, C. a., Repperger, D. W., Kinsler, R., Bharwani, G., & Kender, D. (2007). A quantitative model of the human-machine interaction and multi-task performance: A strategy function and the unity model paradigm. *Computers in Biology and Medicine*, *37*(9), 1259–1271.

Phillips, C. A., Kinsler, R. E., Repperger, D. W., Mandal, J., Neidhard-Doll, A. T., & Kender, D. M. (2013). A human–machine interaction strategy function: information throughput and weighting with application to Multiple-Attribute-Task-Battery. *Theoretical Issues in Ergonomics Science*, *14*(4), 379–401.

Redick, T. S., & Engle, R. W. (2006). Working memory capacity and attention network test performance. *Applied Cognitive Psychology*, *20*(5), 713–721.

Repperger, D. W., Phillips, C. A., & Chelette, T. L. (1995). a Study on Spatially Induced Virtual Force With an Information-Theoretic Investigation of Human-Performance. *Ieee Transactions on Systems Man and Cybernetics*, *25*(10), 1392–1404.

Rice, J. (2006). *Mathematical statistics and data analysis*. Nelson Education.

Rieck, A. M., Ogden, G. D., & Anderson, N. S. (1980). An investigation of varying amounts of component-task practice on dual-task performance. *Human Factors*, 22(3), 373-384.

Roscoe, A. H. (1992). Assessing pilot workload. Why measure heart rate, HRV and respiration?. *Biological psychology*, 34(2), 259-287.

Roscoe, A. H. (1993). Heart rate as a psychophysiological measure for in-flight workload assessment. *Ergonomics*, 36(9), 1055-1062.

Rupp, M. A., Oppold, P., & McConnell, D. S. (2014). Evaluating input device usability as a function of task difficulty in a tracking task. *Ergonomics*, (February), 1–14.

Ruthruff, E., Johnston, J. C., & Van Selst, M. (2001). Why practice reduces dual-task interference. *Journal of Experimental Psychology. Human Perception and Performance*, *27*(1), 3–21.

Ruthruff, E., Johnston, J. C., Van Selst, M., Whitsell, S., & Remington, R. (2003). Vanishing dual-task interference after practice: has the bottleneck been eliminated or is it merely latent? *Journal of Experimental Psychology. Human Perception and Performance*, *29*(2), 280–289.

Salomon, K. A., Ferraro, F. R., Petros, T., Bernhardt, K., & Rhyner, K. (2015). Individual Differences in Multitasking Performance. *59th Annual Meeting of the Human Factors and Ergonomics Society*, 887–891.

Salthouse, T. A., & Miles, J. D. (2002). Aging and time-sharing aspects of executive control. *Memory & Cognition*, 30(4), 572-582.

Salvucci, D. D., & Taatgen, N. A. (2011). Toward a unified view of cognitive control. *Topics in cognitive science*, 3(2), 227-230.

Sanchez-Cubillo, I., Perianez, J. A., Adrover-Roig, D., Rodriguez-Sanchez, J. M., Rios-Lago, M., Tirapu, J. E. E. A., & Barcelo, F. (2009). Construct validity of the Trail Making Test: role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *Journal of the International Neuropsychological Society*, 15(3), 438.

Santiago-espada, Y., Myer, R. R., Latorella, K. A., & Comstock, J. R. (2011). *The Multi-Attribute Task Battery II (MATB-II) Software for Human Performance and Workload Research : A User ' s Guide NASA/TM–2011-217164*.

Schneider, W., & Detweiler, M. (1988). The role of practice in dual-task performance: toward workload modeling in a connectionist/control architecture. *Human Factors*, *30*(5), 539–566.

Schneider, W. (1985). Training high performance skills: Fallacies and guidelines. *Human Factors*, *27*(3), 285–300.

Seow, S. C. (2005). Information Theoretic Models of HCI: A Comparison of the Hick-Hyman Law and Fitts' Law. *Human-Computer Interaction*, *20*(3), 315–352.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System's Technical Journal*, *27*(I), 379–423.

Shannon, C. E., & Weaver, W. (1949). The mathematical theory of information.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4), 591-611.

Sigman, M., & Dehaene, S. (2008). Brain mechanisms of serial and parallel processing during dual-task performance. *Journal of Neuroscience*, 28(30), 7585-7598.

Singh, A. L., Tiwari, T., & Singh, I. L. (2009). Effects of Automation Reliability and Training on Automation- Induced Complacency and Perceived Mental Workload, *35*(October), 9–22.

Singh, A. L., Tiwari, T., & Singh, I. L. (2010). Performance feedback, mental workload and monitoring efficiency. *Journal of the Indian Academy of applied Psychology*, 36(1), 151-158.

Singh, I. L., Sharma, H. O., & Singh, A. L. (2005). Effect of Training on Workload in Flight Simulation Task Performance, *31*(1), 83–91.

Smith, M. E., & Gevins, A. (2005). Neurophysiologic Monitoring of Mental Workload and Fatigue During Operation of a Flight Simulator, *5797*, 116–126.

Society, E., Andrews, S., & Columbia, B. (2016). A Review of Statistical Power Analysis Software, (January 1997).

Sörqvist, P. (2010). High working memory capacity attenuates the deviation effect but not the changing-state effect: Further support for the duplex-mechanism account of auditory distraction. *Memory & Cognition*, *38*(5), 651–658.

Speck-Planche, A., & Cordeiro, M. N. D. S. (2015). Multitasking models for quantitative structure-biological effect relationships: current status and future perspectives to speed up drug discovery. *Expert Opinion on Drug Discovery*, *10*(3), 245–56.

Spink, A., Cole, C., & Waller, M. (2008). Multitasking Behavior, 93–118.

Spink, A., Park, M., Jansen, B. J., & Pedersen, J. (2006). Multitasking during Web search sessions. *Information Processing and Management*, *42*(1 SPEC. ISS), 264–275.

St. John, M., Harris, W. C., & Osga, G. (1997, October). Designing for multi-tasking environments: Multiple monitors vs. multiple windows. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 41, No. 2, pp. 1313-1317). Sage CA: Los Angeles, CA: SAGE Publications.

Stafford, T., & Dewar, M. (2014). Tracing the trajectory of skill learning with a very large sample of online game players. *Psychological Science*, *25*(2), 511–8.

Steichen, B., Wu, M. M., Toker, D., Conati, C., & Carenini, G. (2014, July). Te, Te, Hi, Hi: Eye gaze sequence analysis for informing user-adaptive information visualizations. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 183-194). Springer International Publishing.

Stern, J. A., & Skelly, J. J. (1984, October). The eye blink and workload considerations. In *Proceedings of the Human Factors Society Annual Meeting* (Vol. 28, No. 11, pp. 942-944). Sage CA: Los Angeles, CA: Sage Publications.

Stern, J. A., Walrath, L. C., & Goldstein, R. (1984). The endogenous eyeblink. *Psychophysiology*, 21(1), 22-33.

Sutter, C., Oehl, M., & Armbrüster, C. (2011). Practice and carryover effects when using small interaction devices. *Applied Ergonomics*, *42*(3), 437–444.

Tokuda, S., Palmer, E., Merkle, E., & Chaparro, A. (2009, October). Using saccadic intrusions to quantify mental workload. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 53, No. 12, pp. 809-813). Sage CA: Los Angeles, CA: SAGE Publications.

Turley-Ames, K. J., & Whitfield, M. M. (2003). Strategy training and working memory task performance. *Journal of Memory and Language*, 49(4), 446-468.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent?. *Journal of memory and language*, 28(2), 127-154.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505.

Van Orden, K. F., Limbert, W., Makeig, S., & Jung, T. P. (2001). Eye activity correlates of workload during a visuospatial memory task. *Human Factors*, *43*(1), 111–121.

Van Winsun, W., Sergeant, J., & Geuze, R. (1984). The functional significance of event-related desynchronization of alpha rhythm in attentional and activating tasks. *Electroencephalography and Clinical Neurophysiology*, 58(6), 519-524.

Veltman, J. A., & Gaillard, A. W. K. (1996). Physiological indices of workload in a simulated flight task. *Biological psychology*, 42(3), 323-342.

Veltman, J. A., & Gaillard, A. W. K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics*, 41(5), 656-669.

Verwey, W. B. (1996). Buffer Loading and Chunking in Sequential Keypressing, *22*(3), 544–562.

Walters, C. M. (2012). Application of the human-machine interaction model to Multiple Attribute Task Battery (MATB): Task component interaction and the strategy paradigm.

Wang, Y. (2009). Analyses on limitations of information theory. *2009 International Conference on Artificial Intelligence and Computational Intelligence, AICI 2009*, *1*, 85–88.

Watson, J. M., Bunting, M. F., Poole, B. J., & Conway, A. R. A. (2005). Individual differences in susceptibility to false memory in the Deese-Roediger-McDermott paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 76–85.

Webster, M. (2017). Merriam-Webster online dictionary.

Wickens, C. D. (1991). Processing resources and attention. *Multiple-task performance*, 1991, 3-34.

Wickens, C. D. (2008). Multiple Resources and Mental Workload, *50*(3), 449–455.

Wickens, C. D., Dixon, S. R., & Ambinder, M. S. (2006). Workload and automation reliability in unmanned air vehicles. In *Human factors of remotely operated vehicles* (pp. 209-222). Emerald Group Publishing Limited.

Wickens, C. D., Gutzwiller, R. S., & Santamaria, A. (2015). Discrete task switching in overload: A meta-analyses and a model. *International Journal of Human Computer Studies*, *79*, 79–84.

Wickens, C. D., Mountford, S. J., & Schreiner, W. (1981). Multiple resources, task-hemispheric integrity, and individual differences in time-sharing. *Human factors*, 23(2), 211-229.

Wilson, G. F. (1992). Applied use of cardiac and respiration measures: Practical considerations and precautions. *Biological Psychology*, 34(2), 163-178.

Wilson, G. F. (1993). Air-to-ground training missions: a psychophysiological workload analysis. *Ergonomics*, 36(9), 1071-1087.

Wilson, G. F., Lambert, J. D., & Russell, C. A. (2000, July). Performance enhancement with real-time physiologically controlled adaptive aiding. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 44, No. 13, pp. 61-64). Sage CA: Los Angeles, CA: SAGE Publications.

Zhang, Y., & Hornof, A. J. (2014). Understanding multitasking through parallelized strategy exploration and individualized cognitive modeling. *CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3885–3894.

Zimmermann, P., & Fimm, B. (2002). A test battery for attentional performance. Applied neuropsychology of attention. *Theory, diagnosis and rehabilitation*, 110-151.

APPENDICES

Appendix A: Demographic Questionnaire

Participant ID: _____

Gender: ____Male ____Female

Age: _____

Handedness: ____ Right ____ Left

Native/Primary language: ____ English ____ Other

1. Are you in overall good health?

2. Do you have any previous experience with MATB-II or similar systems?

Appendix B: Handedness Survey

Please indicate with a one (1) your preference in using your left or right hand in the following tasks. Where the preference is so strong you would never use the other hand, unless absolutely forced to, put a two (2). If you are indifferent, put a one in each column (1 | 1).

Some of the activities require both hands. In these cases, the part of the task or object for which hand preference is wanted is indicated in parentheses.

Task / Object	Left Hand	Right Hand
1. Writing		
2. Drawing		
3. Throwing		
4. Scissors		
5. Toothbrush		
6. Knife (without fork)		
7. Spoon		
8. Broom (upper hand)		
9. Striking a Match (match)		
10. Opening a Box (lid)		
Total checks:	LH =	RH =
Cumulative Total	CT = LH + RH =	
Difference	D = RH – LH =	
Result	R = (D / CT) × 100 =	
Interpretation: (Left Handed: R < -40) (Ambidextrous: -40 ≤ R ≤ +40) (Right Handed: R > +40)		

Appendix C: Instruction of Automated OSPAN Task

The automated OSPAN task includes items (letters) to remember and a distracting activity in the form of math problem solving (Unsworth et al., 2005). There are three practice sessions (letter span, math problem, both of them combined) and one experimental session.

1. Practice Session I

   Letters appear one at a time on the screen and you are required to remember the letters in the presented order. At recall, you will see a 4 X 3 matrix of letters (F, H, J, K, L, N, P, Q, R, S, T, and Y). After recall, the program provides feedback about the number of letters correctly recalled in the current set.

2. Practice Session II

   This session involves simple mathematical problem solving. You need to solve several True/False math problems (e.g., (1*2) + 1= 3 True or False). During the task, the computer will calculate the mean time required to solve the equations for you.

3. Practice Session III

   During this session, you need to perform both the letter recall and math portions together. First, a letter will be presented. After a few seconds the letter disappears and a math problem is presented. The math problem is then replaced with a "True/False" prompt. Once the you responds "True" or "False", another letter appears on the screen and the sequence is repeated.

4. OSPAN Test Trial

After completion of practice sessions, the program will progress to the real trials similar to the final session of practice set. The experimental trial consists of 75 letters and 75 math problems, broken down into smaller segments with a maximum segment length of 7 sets. There are 75 trials in the experimental session and score range of this test is 0 – 75.

At the end of the task, the program will report five scores: OSPAN score, total number correct, math errors, speed errors, and accuracy errors. The OSPAN score, uses the traditional absolute scoring method: the sum of all perfectly recalled sets. The second score, "total number correct," is the total number of letters recalled in the correct position. Three types of errors are reported. "Math errors" represent the total number of task errors, which is then broken down into "speed errors" and "accuracy errors". "Speed errors" quantify the amount of times participants run out of time in attempting to solve a given math operation. "Accuracy errors" represent the number of incorrectly solved math operations.

Appendix D: WRS Instruction

Mental Demand

How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc)? Was the task easy or demanding, simple or complex, exacting or forgiving?

Physical Demand

How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Temporal Demand

How much time pressure did you feel due to the rate of pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

Performance

How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

Effort

How hard did you have to work (mentally and physically) to accomplish your level of performance?

Frustration

How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Appendix E: NASA-TLX Pairwise Comparisons of Factors

Instructions: select the member of each pair that provided the most significant source of workload variation in these tasks.

PD	MD
TD	MD
PE	MD
FR	MD
EF	MD
TD	PD
PE	PD
FR	PD
EF	PD
TD	PE
TD	FR
TD	EF
PE	FR
PE	EF
EF	FR

Note:

MD	mental demand
PD	physical demand
TD	temporal demand
PE	performance
EF	effort
FR	frustration

For Experimenter Use Only

Tally of Importance Selections	
MD	
PD	
TD	
PE	
FR	
EF	
Sum	15

Appendix F: Summary of User Performance

	Phase I			Phase II		
Task Difficulty	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**
RT_Light	2.28 (0.64)	2.70 (0.81)	2.59 (0.62)	2.11 (0.64)	2.33 (0.72)	2.08 (0.48)
RT_Scale	4.01 (1.39)	4.40 (1.43)	5.25 (1.36)	3.48 (1.59)	3.95 (1.28)	4.30 (1.43)
RT_COMM	1.55 (0.53)	1.71 (0.55)	2.18 (0.46)	1.58 (0.66)	1.62 (0.60)	2.19 (0.51)
RR_Light	0.99 (0.02)	0.99 (0.02)	0.92 (0.05)	0.99 (0.05)	1.00 (0.01)	0.96 (0.04)
RR_Scale	0.82 (0.28)	0.85 (0.23)	0.88 (0.16)	0.80 (0.30)	0.85 (0.26)	0.89 (0.16)
RR_COMM	0.93 (0.08)	0.93 (0.08)	0.85 (0.11)	0.95 (0.09)	0.92 (0.10)	0.90 (0.09)
RR_RMAN	0.89 (0.09)	0.89 (0.08)	0.86 (0.10)	0.92 (0.07)	0.90 (0.07)	0.88 (0.09)
RMSD	18.93 (7.95)	30.80 (12.03)	47.90 (19.61)	17.32 (6.23)	27.35 (8.26)	39.33 (11.34)
WRS	40.29 (18.69)	49.54 (17.16)	65.42 (15.84)	39.20 (20.15)	46.07 (17.23)	60.87 (16.26)
%BKD	-12.31% (11.17%)	-14.01% (11.94%)	-10.92% (21.98%)	-4.81% (24.23%)	-7.63% (22.69%)	-13.61% (15.09%)
%BKR	-44.20% (32.97%)	-44.36% (25.79%)	-52.83% (32.29%)	-37.27% (36.70%)	-43.95% (32.80%)	-47.65% (37.16%)

Appendix G: Summary of Tukey's HSD Test Results

	Task Difficulty		p-Value
RT_Light	Medium	Low	0.0022*
	Medium	High	0.1335
	High	Low	0.3056
RT_Scale	High	Low	<.0001**
	High	Medium	0.0101*
	Medium	Low	0.0909
RT_COMM	High	Low	<.0001**
	High	Medium	<.0001**
	Medium	Low	0.3986
RR_Light	Medium	High	<.0001**
	Low	High	<.0001**
	Medium	Low	0.9975
RR_COMM	Low	High	<.0001**
	Medium	High	0.0004*
	Low	Medium	0.6486
RR_RMAN	Low	High	0.0275*
	Medium	High	0.1106
	Low	Medium	0.8356
RMSD	High	Low	<.0001**
	High	Medium	<.0001**
	Medium	Low	<.0001**
WRS	High	Low	<.0001**
	High	Medium	<.0001**
	Medium	Low	0.0038*

Note. * significance level: $p < 0.05$. ** significance level: $p < 0.0001$.