Toward Enhancing Multitasking Performance: Modeling, Prediction, and Intervention Design

Jiaxin Li

A proposed dissertationsubmitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2025

Reading Committee:

Ji-Eun Kim, Chair

Linda Ng Boyle, Chair

Prashanth Rajivan

Program Authorized to Offer Degree:
Industrial and Systems Engineering

University of Washington

Abstract

Toward Enhancing Multitasking Performance: Modeling, Prediction, and Intervention Design

Jiaxin Li

Co-Chairs of the Supervisory Committee:
Ji-Eun Kim
Industrial & Systems Engineering

Linda Ng Boyle
Industrial & Systems Engineering and Civil & Environmental Engineering

Multitasking involves performing more than one task in parallel or in a serial manner. Improving multitasking performance is particularly crucial in safety-critical environments, where performance decrement can lead to serious or even life-threatening consequences. For example, engaging in secondary tasks while driving, commonly known as distracted driving, is one of the main contributors to traffic accidents. In the field of aviation, several fatal aircraft crashes have also been linked to pilots' errors during multitasking. Despite the crucial need to enhance multitasking performance, existing studies continue to rely on retrospective behavioral measurements, which are insufficient for continuously tracking and predicting individuals' multitasking performance. Furthermore, despite the widespread use of automation in human-system interactions, there is a lack of research on how to design automation in multitasking environments.

This dissertation aims to answer four research questions: 1) Which factors impact multitasking performance in practical scenarios? 2) Which neurophysiological responses indicate changes in multitasking performance? 3) How can multitasking performance be estimated over time using probabilistic modeling? 4) How multitasking performance can be enhanced by automation? To answer these research questions, a mix of survey, behavioral, and neurophysiological data recorded from both controlled experiments and field studies was used.

The findings of this dissertation can be applied in safety-critical settings to reduce operators' multitasking errors, enable timely and effective interventions, and ultimately mitigate safety risks.

TABLE OF CONTENTS

LIST OF FIGURES

GLOSSARY

AIC: Akaike Information Criterion

AOI: area of interest

ANOVA: analysis of variance

ART: aligned rank transform

ASW: average silhouette width

BIC: Bayesian Information Criterion

BN: Bayesian network

DBN: dynamic Bayesian network

DNN: deep neural networks

EEG: electroencephalogram

GSR: galvanic skin response

HRV: heart rate variability

IES: inverse efficiency score

LR: linear regression

MATB-II: the Multi-Attribute Task Battery II

MSE: mean squared error

MRT: Multiple Resource Theory

NASA: the National Aeronautics and Space Administration

NASA-TLX: NASA Task Load Index

RF: random forests

RQ: research question

SGE: stationary gaze entropy

SHRP2: the second Strategic Highway Research Program

WFH: work from home

ACKNOWLEDGMENTS

I would like to express my gratitude to everyone who has supported me throughout my Ph.D. journey.

I thank my advisor, Dr. Ji-Eun Kim and Dr. Linda Ng Boyle, for your guidance and support. The knowledge I have learned from you and your mentorship are invaluable. You have been role models to me over the years.

To my committee members, Dr. Prashanth Rajivan and Dr. Andrea Stocco, thank you for your insightful comments and feedback, which have helped improve the quality of my research.

I would like to extend my appreciation to my lab mates from the HAS Lab and the HFSM Lab at UW, David Prendez, Veronika Kettel, Elizabeth A. Higgins, Zishu Ling, Daisy Xiao, Hsuan-ching Wang, Grace Douglas, Sami Park, and Mayuree Binjolkar. I am lucky to have had your companionship both in work and in life.

I am also grateful to my friends, Tianchen Sun, Pariyakorn Maneekul, Auur Anna Jónsdóttir, Lun Li, You Chen, Hao Tang, Yilun Xing, Xiaonan Sun, and Annie Huang, for your constant friendship over the years.

To my boy friend, Daniel Xu, thank you for standing by me throughout the time. You have given me the strength and encouragement I needed to overcome the challenges along the way.

Lastly, I want to express my deepest gratitude to my dear parents, Hong Yu and Huaihong Li, for your unconditional love and unwavering support. I owe this achievement to everything you have given me.

Chapter 1

INTRODUCTION

Multitasking is a common part of our daily life. People often choose to multitask to get things done more quickly, but sometimes multitasking might lead to the opposite outcome and hurt our performance. Performing multitasking in a timely and accurate manner has become an essential and valuable skill in modern workplaces (Chérif et al., 2018; Liu and Nam, 2018). Advancements in technology and media have increased the volume of information individuals need to handle, which also brings the need to navigate multiple devices and tools simultaneously. However, multitasking can sometimes result in reduced performance for each sub-task. For example, replying to emails during a conference call may cause you to miss important parts of the discussion. Given the prevalence of multitasking, it is important to mitigate the potential downsides of multitasking and enhance multitasking performance (Chérif et al., 2018).

In safety-critical environments, multitasking errors can lead to significant consequences. In clinical settings, there is a growing demand for clinicians to multitask to ensure efficient healthcare treatment (Heng, 2014; Weigl et al., 2013). Poor multitasking has been associated with medical errors and infection control failures, posing a severe risk to patient safety (Weigl et al., 2013). A study by Westbrook et al. (2018) found higher prescribing error rates when emergency physicians multi-tasked by placing medication orders and handling other tasks at the same time. Another study based on the Manufacturer and User Facility Device Experience database revealed that 48% of reported fatalities associated with infusion pumps were due to nurses' errors during multitasking (Clark et al., 2006). In aviation, air-traffic controllers are responsible for managing multiple tasks simultaneously, such as managing aircraft and air-traffic flow, giving accurate commands to pilots, and responding to any weather change or emergencies (Bernhardt et al., 2019; Hopkin, 2017). A non-negligible proportion of aviation accidents and incidents can be attributed to controllers' multitasking errors (Pape et al., 2001). Beyond the workplace, driving is one of the most common multitasking situations in our daily lives. The prevalent use of smartphones and in-car entertainment systems has promoted more multitasking during driving, which is a major contributor to vehicle-pedestrian collisions (Regan et al., 2011). The National Highway Traffic Safety Administration reported that distracted driving contributed to 23,000 fatalities in traffic accidents from 2012 to 2018 (Sajid Hasan et al., 2022).

The growing demand for multitasking, along with the potential for severe consequences of multitasking errors, highlights the need to explore effective strategies for enhancing multitasking performance. Two complementary approaches could be used to tackle this challenge. One approach is to develop robust models that accurately predict multitasking performance, enabling timely intervention and reducing the impact of errors. The second approach is to design effective interventions that improve operators' ability to handle multitasking.

Chapter 2

LITERATURE REVIEWS

This dissertation aims to enhance human operators' multitasking performance in practical scenarios, thereby improving the overall productivity and safety in high-risk environments. This chapter provides an overview of relevant studies to establish a deeper understanding of the topic, covering theories that explain performance decrements in multitasking, key factors and reliable indicators of multitasking performance, appropriate modeling techniques, and intervention methods. The chapter concludes with a discussion of the research goals and questions that will be addressed in the subsequent chapters.

2.1 Multitasking Paradigms

Multitasking refers to performing two or more tasks either simultaneously or sequentially (Chérif et al., 2018; Spink et al., 2008). There are two main paradigms of multitasking: dual-task and task switching (Różańska and Gruszka, 2020; Spink et al., 2008). In practical multitasking environments, individuals might also process multiple tasks in a combination of dual-task and task-switching (Broeker et al., 2018).

2.1.1 Dual-task

The dual-task paradigm involves performing two tasks in parallel (Himi et al., 2019; Lin, 2013). Some studies also extend the concept of "dual-task" to include conducting more than two tasks simultaneously (Chérif et al., 2018; Lui and Wong, 2020). Individuals' attention is divided in dual-tasking, and their performance is diminished compared to when each task is performed in isolation (Koch et al., 2018). This decrement is termed dual-task interference. For example, when two stimuli are presented with varying time overlap, the reaction time for the second stimulus becomes longer as the intervals between the onset of the two stimuli become shorter (Pashler, 1994).

2.1.2 Task Switching

Task switching refers to individuals shifting among multiple tasks and processing them in a sequential manner (Koch et al., 2018; Monsell, 2003). In situations where dual-tasking is impractical, task switching occurs, requiring individuals to engage in only one task at a time and temporarily suspend the others (Lee et al., 2008). Researchers have observed a switch cost after task switching: operators need a period to initiate the upcoming task and recover the efficiency, leading to a performance decrement (e.g., longer response times or higher error rates) compared to non-switching or task-repetition conditions (Broeker et al., 2018; Monsell, 2003; Spink et al., 2008).

2.2 Performance Decrement During Multitasking

Multitasking often leads to reduced performance compared to focusing on each task separately. This section introduces theories and explanations for the performance decrement observed in multitasking within the dual-task and task switching paradigms. These theories not only provide a foundation for modeling multitasking performance but also offer theoretical insights into how such performance decrements can be mitigated.

2.2.1 Dual-task Interference

Since the last century, researchers have developed diverse theories to explain the multitasking performance decrement in dual-task scenarios. The Bottleneck Theory posits that specific stages in information processing act as bottlenecks, restricting the ability to process multiple tasks simultaneously (Pashler, 1993). This assumption helps explain why individuals face challenges in dual-tasking compared to performing a single-task. The Resource Theory assumes a general resource with limited capacity that can be allocated among multiple tasks and all stages of information processing, offering an explanation for the fluctuation in multitasking performance corresponding to changes in task demand (Kahneman, 1973; Navon and Gopher, 1979; Wickens, 2008). However, these theories have limitations as they fall short of explaining practical situations where multiple demanding tasks can be executed in a perfect time-sharing manner (Meyer and Kieras, 1997; Pew and Mavor, 1998; Wickens, 2020).

Multiple Resource Theory (MRT) (Wickens, 2002) explains the performance decrement in dual-task by hypothesizing that an individual's attentional resources can be categorized into multiple resource pools. Tasks requiring resources from the same pool are more likely to interfere with each other, whereas tasks that require resources from different pools within each dimension are less prone to interfere. The interference between tasks can result in performance decrements. MRT is based on the Resource Theory but discards the assumption of a central unitary resource and, instead, introduces dimensions representing distinct resource pools (Kahneman, 1973; Navon and Gopher, 1979; Wickens, 2008). The resource pools in MRT include the stages of information processing (perceptual/cognitive or action), the codes of processing (spatial or verbal), and modalities (auditory or visual) (Wickens, 2002). The dichotomies in each dimension are considered separate resource pools. MRT explains how humans manage and execute multiple tasks and accommodates the situations where certain tasks can be performed simultaneously without significant interference (Wickens, 2008,2).

- *Stages of information processing.* The dimension of information processing stage assumes two separate pools for perceptual/cognitive activities and execution of responses. In practical terms, this suggests that the interference caused by responding to one task is less likely to affect another task that is more perceptual or cognitively demanding. This division aligns with the evidence that perceptual-cognitive processing is linked to the reticular activating system, whereas the response stage is associated with the limbic system (Meyer and Kieras, 1997).

- *Codes of information processing.* The processing code dimension refers to spatial and verbal information being processed using separate resource pools. This dichotomy can be explained by the fact that the processing of spatial information is primarily associated with the right hemisphere of the brain, while verbal information processing is typically associated with the left hemisphere (Nagel et al., 2013).

- *Modalities.* In the modality dimension, MRT proposes distinct pools for auditory and visual processing. For instance, listening to music is less likely to interfere with reading news on your cell phone, while reading news during driving is more likely to interfere because both tasks demand the resource of visual perception. This division aligns with the evidence that visual or auditory stimuli activate different lobes of the brain (O'Leary et al., 1997).

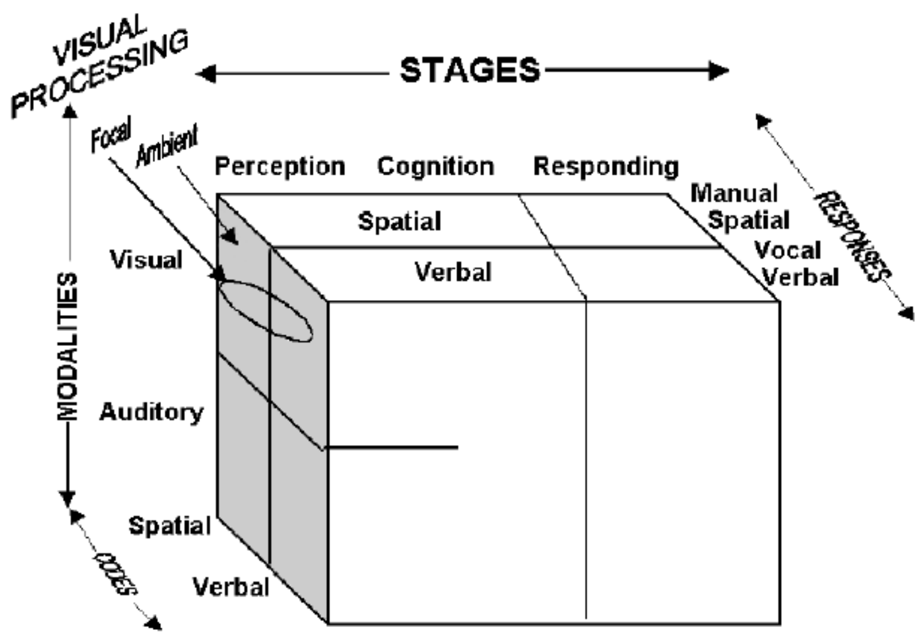


Figure 2.1: The Multiple Resource Theory model (Adapted from Wickens (2002)).

MRT has been employed to explain multitasking performance in diverse environments. For instance, it served as the theoretical foundation for the development of an air traffic controller behavioral model (Volf et al., 2014). This model was constructed based on the assumption of limited pools of modalities and stages of information processing to simulate the workload change of air traffic controllers. In driving scenarios, countermeasures have been implemented to discourage smartphone use while driving. According to MRT, both activities are visually demanding, and using a smartphone is highly likely to interfere with driving, posing safety risks (Hassani et al., 2017).

2.2.2 Task Switching Cost

Previous studies have identified two main explanations for the switching cost (Evans et al., 2015). The first explanation for switching cost is task-set reconfiguration. A task-set refers to the mental set-up that enabling an individual to perform a task effectively. The task-set reconfiguration refers to the process that individuals prepare for the new coming task, including shifting attention to the new task and retrieving information about what to do and how to do it from working memory (Courage et al., 2015; Kiesel et al., 2010; Monsell, 2003). This process requires a period of time and leads to a longer reaction time. The second reason for switching cost is task-set inertia. Task-set inertia describes the interference from the previous task, where the task-set associated with a previous task persists and influences an individual's performance on the new task (Courage et al., 2015; Evans et al., 2015; Monsell, 2003). It results in a longer reaction time or a higher likelihood of errors during and even after task-switching.

2.3 Factors Impacting Multitasking Performance

Identifying the factors that impact multitasking performance is essential for modeling, predicting, and enhancing it. A clear understanding of these factors offers insight into how individuals are influenced by multitasking environments (Chérif et al., 2018). These factors can be broadly categorized into task-related factors and individual factors. This section reviews several key factors that are further examined in the following chapters.

2.3.1 Task Complexity

Task complexity has been shown to affect individuals' performance during multitasking (Liu et al., 2016; Valéry et al., 2019). Task complexity can be defined in several ways, and this dissertation adopts the definition provided by Robinson (2001), which characterizes task complexity as "a result of resource requirement" (Liu and Li, 2012). The higher the information load generated by the task, the greater the task complexity, leading to a greater demand for attentional resources (Liu and Li, 2012; Robinson, 2001). An increase in task complexity is linked to greater interference among tasks in dual-tasking (Wickens, 2020) and is also associated with higher switching costs in task switching (Kiesel et al., 2010;

Monsell, 2003).

2.3.2 Task Repetition

Task repetition refers to the repeated execution of the same tasks. Previous studies have suggested that task repetition makes the tasks more predictable and decreases their demand on attentional resources (Broeker et al., 2018; Kida et al., 2012). Task repetition can also improve attention allocation by directing operators' attention to more task-relevant information or the most beneficial tasks (Griffiths et al., 2004; Lemonnier et al., 2020; Martens and Fox, 2007).

2.3.3 Sensory Modality

Sensory modality refers to the channel through which individuals receive information from environmental stimuli. These stimuli are typically classified based on the sensory receptors they activate (Auvray and Spence, 2008). Typical sensory modalities include visual, auditory, and tactile, but visual and auditory modalities are predominant in most multitasking settings (Hutmacher, 2019). In dual-tasking, according to multiple resource theory (Wickens, 2002), multitasking interference is greater when tasks use the same sensory modality (e.g., two visual tasks) and lower when tasks use different modalities (e.g., one visual task and one auditory task). Similarly, Wahn et al. (2017) demonstrated in a dual-task paradigm that distributing tasks across different sensory channels leads to improved overall performance and better management of individual subtasks. Modality plays a complex role in task-switching, though findings across studies are not entirely consistent (Peng et al., 2018). Some early studies suggest that shifting attention between modalities can lead to performance decrement, resulting in longer response times (Hunt and Kingstone, 2004; Waszak et al., 2003). In contrast, others reported a facilitative effect, where switching between modalities actually reduced task-switching costs, possibly due to interference reduction (Murray et al., 2009). Additionally, some studies suggest that the effect of modality on multitasking performance depends on the interplay between sensory modality and the type of responses (Fintor et al., 2018; Stephan and Koch, 2011).

2.3.4 Task Priority

The priority of a task refers to its level of importance compared to other tasks (Barg-Walkow and Rogers, 2017). The priority level assigned to a task influences the amount of attentional resources it is allocated (Iani and Wickens, 2007; Valéry et al., 2019). A task assigned higher priority receives a larger share of the available resources when competing with other tasks. In task-switching scenarios, individuals are more inclined to switch to prioritized tasks or stay on the ongoing task if it has the highest priority (Wickens et al., 2015).

2.3.5 Individual Factors

There are individual differences in multitasking performance. Aging can significantly impact individuals' multitasking ability, as it has been related to declines in information processing speed and attention capability (McAvinue et al., 2012; Verhaeghen et al., 2003). Previous studies also demonstrate that older people have reduced working memory capacity and deteriorated executive functions, which further limit their ability to multitask. As for scenario-specific multitasking, such as distracted driving, individuals' expertise level can also affect their performance (Horrey et al., 2008). Studies have shown that novice drivers are more susceptible to the detrimental effects of multitasking while driving, leading to a higher likelihood of accidents (Klauer et al., 2014; Stavrinos et al., 2020).

Despite theoretical and empirical findings on key factors affecting multitasking, many existing studies are limited in their ability to translate these insights into practical scenarios. Most studies on multitasking have focused on highly controlled experiments, where conditions are regulated to isolate specific factors. To conduct a more comprehensive analysis of influential factors in multitasking, this research addresses the following question using two datasets: one from a simulator replicating pilots' routine tasks and another from a naturalistic driving study:

RQ1: Which factors impact multitasking performance in practical scenarios?

2.4 Neurophysiolgoical Indicators of Multitasking Performance

Continuous monitoring of multitasking performance is essential for accurate modeling and timely intervention, especially in safety-critical environments. Performance decrement in multitasking fluctuates over time based on task demands, interactions between tasks, and operators' strategies (Plummer and Eskes, 2015). Recent advancements have enhanced the portability, unobtrusiveness, and automated signal processing of neurophysiological sensors (Dehais et al., 2020). These improvements have made them highly effective for detecting changes in multitasking performance across diverse task settings.

2.4.1 Eye Movements

An eye tracker detects the position and movement of the eyes using infrared light or high-resolution cameras (Kredel et al., 2017). Fixation metrics are commonly used components of eye movement in multitasking research. Eye fixation occurs when a person's eyes maintain a steady gaze on a point for at least 100 milliseconds (Borys and Plechawska-Wójcik, 2017; Castet and Crossland, 2012). The number of fixations and fixation duration are key eye-tracking metrics in multitasking studies, reflecting the engagement of visual attention and the time required to process stimuli on a visual target (Eckstein et al., 2017). Areas of Interest (AOI) metrics are used to examine task-specific eye movement patterns and evaluate how participants allocate their attention during multitasking (Becker, 2011; Cui et al., 2024). Analyzing fixation and AOI metrics provides insights into how individuals allocate their attention during multitasking and reveals the underlying cognitive processes and strategies.

2.4.2 Brain Signals

An electroencephalogram (EEG) captures the brain's electrical activity through electrodes placed on the scalp (Jackson and Bolger, 2014). Previous studies often utilize the power spectral density of frequency bands to understand how the brain functions and adapts to task designs (Klimesch, 2018). Frequency band power is derived by decomposing the EEG signals into several frequency bands, including delta (1–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (30–100 Hz) (Kim and Im, 2018). Among them, theta- and alpha-band power are associated with attention allocation. Theta-band power can reflect the activation of high-level cognitive processes and is sensitive to the task demand (Cavanagh and Frank, 2014; Gevins et al., 1998; Klimesch, 1999). Alpha-band power is related to processing task-related information and can also indicate the demand for attention during multitasking (Foxe and Snyder, 2011; Klimesch, 2018; Puma et al., 2018).

2.4.3 Multisensory Data

Rather than solely relying on one measurement, recent studies have been using multisensor data to model human performance and cognitive states. Multisensor approaches are favored because they provide complementary data that single sensors might miss, mitigate the noise susceptibility of single sensors, and ultimately improve prediction accuracy (Al Imran et al., 2024; Brunzini et al., 2024; Iqbal et al., 2024). For example, Al Imran et al. (2024) argue that single-sensor approaches may fail to capture certain aspects of fatigue, while multisensor methods can capture complementary information to enhance detection accuracy. Khosravi et al. (2024) combined eye-tracking, electrodermal activity, and photoplethysmography to monitor mind-wandering in students. Their integrated model showed higher prediction accuracy than those relying on only one sensor by obtaining higher prediction accuracy.

Although research has indicated the potential of neurophysiological responses in reflecting multitasking performance across diverse task settings, there is a lack of studies using them together to continuously monitor multitasking performance. Most studies still rely on measures such as overall error rates or averaged reaction times for post-task assessment of multitasking performance (Meng et al., 2021; Szumowska and Kossowska, 2016; Westbrook et al., 2018), which are asynchronous and lack information during the multitasking process (Moacdieh et al., 2020). To address these limitations, I propose the following research question in this dissertation:

RQ2: Which neurophysiological responses indicate the changes in multitasking performance?

2.5 Probabilistic Modeling

One of the main challenges in using neurophysiological responses in real-world settings is their low signal-to-noise ratio. These responses are influenced by various sources of environmental uncertainty. Probabilistic models can account for these uncertainties by identifying the probabilistic distribution of variables (Ghahramani, 2015; Murphy et al., 2002). Probabilistic models allow researchers to incorporate prior knowledge into the modeling process.

Leveraging established theoretical and empirical findings as priors can enhance the model's interpretability and generalizability (Ghahramani, 2013). Given these advantages, the application of probabilistic modeling in multitasking research has not yet been fully explored. The potential of probabilistic models to estimate multitasking performance over time remains an open question, leading to the third research question in this dissertation:

RQ3: How can multitasking performance be predicted over time using probabilistic modeling?

2.6 Automation Intervention

Automation is recognized as an effective tool for enhancing multitasking performance by reducing the demand for attentional resources across various stages of information processing (Onnasch et al., 2014; Parasuraman and Wickens, 2017; Sato et al., 2023b). Automation is defined as a system that performs partial or full functions previously handled by by human operators (Chérif et al., 2018; Parasuraman et al., 2000). It has been widely applied in real-world settings to assist human operators in multitasking, particularly in safety-critical areas. An example of automation in driving is advanced driver assistance systems, which are now commonly equipped in modern vehicles. These systems improve driving performance by assisting drivers with various manipulations such as blind-spot detection, lane change assistance, and cruise control (Nidamanuri et al., 2021).

2.6.1 Full Automation

Full automation refers to systems that execute all tasks without human intervention, thereby eliminating the need for manual control. It has been implemented in some industries, such as automotive manufacturing, where automated assembly lines perform repetitive tasks like welding and significantly improve the productivity Papulová et al. (2022). One major limitation of full automation is that no automation system is 100% reliable; machines may fail or struggle to respond effectively to unpredictable events. Consequently, human monitoring is still necessary for some critical tasks to prevent failures or malfunctions of automation (Ferraro and Mouloua, 2021; Onnasch et al., 2014). Studies have emphasized that monitoring automation can foster over-reliance on automation, which can negatively impact performance and even result in severe consequences (Yamani and Horrey, 2018), or conversely, increase the workload on the human monitor (Endsley, 2017).

2.6.2 Partial Automation

Partial automation has been increasingly integrated into daily activities, allowing systems to handle some tasks while leaving others under human control. For example, partial automation in aviation enhances pilots' multitasking performance by automating routine tasks, thereby reducing total task demands and pilot workload (Wang et al., 2021). The autopilot system maintains altitude, heading, and speed, but pilots remain responsible for communicating with air traffic controllers and handling safety-critical situations, such as those that require the pilots to make evasive maneuvers to avoid collisions (Metzger and Parasuraman, 2017). In addition, pilots are expected to continuously monitor the automated system to detect and mitigate automation failures (Vu et al., 2012).

Compared to full automation, partial automation requires less investment in technology and reduces operator overreliance on automation by keeping humans in the loop (Endsley, 2017). Previous studies indicate that partial automation enables operators to intervene when automated systems encounter errors or human judgment is essential, thereby improving error mitigation and overall safety (Carsten and Martens, 2019; Parasuraman et al., 2000). However, a key challenge in implementing partial automation is determining which tasks should be automated. When designing the automation strategy, system engineers must consider the attributes of each task and the task's impact on human operators, ensuring that the automation supports rather than hinders operators' ability to manage multiple tasks effectively.

Although numerous studies have explored the usefulness of automation in a variety of task settings, these tasks are usually investigated in isolation, and limited effort has been devoted to understanding how to achieve better multitasking performance with the aid of automation (Calhoun, 2022). This limitation of current studies leads to the fourth research question in this dissertation:

RQ4: How multitasking performance can be enhanced by automation?

2.7 Research Goals and Summary of Research Questions

This dissertation aims to develop an integrated, data-driven framework for enhancing multitasking performance by constructing accurate predictive models and exploring effective intervention design. To achieve this objective, four key research questions are addressed, as outlined in Table 2.1.

Table 2.1: Research questions and corresponding chapters.

Research question	**Chapter**
RQ1. Which factors impact multitasking performance in practical scenarios?	Chapters 3 & 4
RQ2. Which neurophysiological responses indicate changes in multitasking performance?	Chapter 4
RQ3. How can multitasking performance be predicted over time using probabilistic modeling?	Chapter 5
RQ4. How multitasking performance can be enhanced by automation?	Chapter 6

To systematically investigate these questions, the dissertation is organized into a series of chapters. Chapter 3 investigated the factors impacting multitasking in a real-world scenario, distracted driving, by analyzing a naturalistic driving dataset; Chapter 4 identified task-related factors of multitasking performance and explored the neurophysiological indicators in a lab experiment; In Chapter 5, I built a dynamic Bayesian Network (DBN) to model individuals' multitasking performance overtime. To validate its effectiveness in modeling multitasking performance, I made comparison analysis with other commonly-used models; In Chapter 6, I conducted a lab experiment examining how to design automation intervention, specifically focusing on how to select tasks for partial automation based on sensory modality and task priority in order to enhance overall multitasking performance.

Chapter 3

MODELING DISTRACTED DRIVING BEHAVIOR BY IDENTIFYING FACTORS FROM NATURALISTIC DATA

Multitasking in driving requires the allocation of limited attentional resources to both driving and non-driving activities. Driving is a unique multitasking situation in which driving is the primary and only critical task, while all other activities are considered secondary tasks. This section seeks to address RQ1 by identifying factors of multitasking performance in a real-world scenario. Dataset from the second Strategic Highway Research Program (SHRP2) were utilized for this study.

This study assessed multitasking performance using four driving performance measures: speed, gas pedal usage, lane position, and the number of hands on the steering wheel. Using k-means clustering, nine secondary tasks were categorized into three groups based on their similarities in these measures. Results from the mix-effects logit model identified the factors contributing to the likelihood of drivers' engagement in the three groups of secondary tasks. This study offers a better understanding of drivers' decision-making in secondary task engagement and provides insights for selecting effective interventions to prevent different types of secondary activities in driving.

3.1 Introduction

Pedestrian-vehicle conflicts are a major concern across the world (WHO, 2018). From 2009 to 2016, pedestrian fatalities at intersections have increased on average 5.2% per year (Hu and Cicchino, 2018). A major contributor of vehicle-pedestrian collisions is driver inattention and driver distraction (Regan et al., 2011; Sajid Hasan et al., 2022). The widespread use of smartphones and in-car entertainment systems have encouraged more non-driving activities during driving. Distracted driving may negatively impact a driver's ability to see pedestrians at intersections.

Multiple resource theory provides a basis for characterizing the secondary tasks with respect to visual, auditory, cognitive, and manual distractions in driving (Kimura et al., 2022; Neyens and Boyle, 2008; Stavrinos et al., 2018; Young et al., 2007). Visual distraction occurs when the drivers' view is on non-roadway objects, and auditory distraction occurs when the driver is focused on an auditory stimulus (e.g., listening to conversation). Cognitive distraction is associated with the state of "mind off the road" (Stavrinos et al., 2018), and manual distraction refers to when the drivers physical movements are not driving related. Visual and manual distractions are often considered to be the greater risks given the visual-manual demanding nature (Young et al., 2007). However, it is common that a secondary task involves two or more distracting components in naturalistic driving scenarios, which makes it challenging to assess the compound effects of multiple distraction components on driving safety (Zhang et al., 2014).

Distracted driving can also be characterized in terms of driving performance measures (Papantoniou et al., 2017; Qin et al., 2019) related to longitudinal and lateral vehicle control (Choudhary and Velaga, 2017a; Kashevnik et al., 2021; Oviedo-Trespalacios et al., 2016), and reaction time to roadway events (Kashevnik et al., 2021; Qin et al., 2019; Svetina, 2016). Given that no single measure can completely capture the impacts of distracted driving, a combination of these performance measures is considered in this study to provide a more comprehensive understanding of the driver (Young et al., 2009).

The goal of this study is to better understand drivers' engagement in secondary tasks, with the aim of providing valuable insights for enhancing pedestrian safety at intersections. To achieve this, the study focused on identifying factors that influence drivers' secondary task behaviors.

3.2 Methods

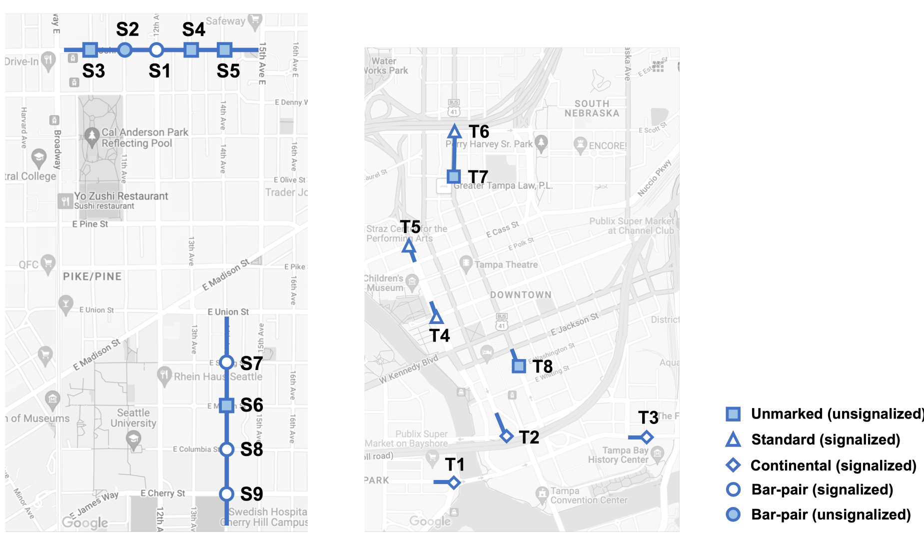
3.2.1 Dataset

SHRP2 is a naturalistic driving study that collected data from 2008 to 2013. More than 3,400 drivers were recruited from six states in the US. The data acquisition system used in the SHRP2 program was capable of collecting time series vehicle and video data from each traversal taken by participants. The recorded videos included the forward view, rear view, face of the driver, and dashboard view using four video cameras (Campbell, 2012).

3.2.2 Site Selection

The study analyzed data from 676 traversals at 17 intersections in Seattle, WA and Tampa, FL. These two cities were selected given their proximity to central business districts and universities, where pedestrian-vehicle interactions were more likely. The 676 traversals came from 406 participants with a range of 1 to 4 traversals per participant (mean=1.67). For each traversal, the data (video and vehicle kinematics) was examined at 50 meters before the crosswalk. Data from 50 meters before the crosswalk was selected to ensure inclusion of drivers recognizing the crosswalk and adjusting their driving behavior accordingly (Pantangi et al., 2021; Varhelyi, 1998).

Table 3.1 and Figure 3.1 show the crosswalk markings and signalization status of each selected intersection. The range of crosswalk markings included unmarked, standard, bar-pair, and continental, as depicted in Figure 3.2. These crosswalk markings were selected as they were readily available in the two cities selected for this study.



(a) Seattle, WA (b) Tampa, FL

Figure 3.1: Selected crosswalks in Seattle and Tampa with map legend.

Table 3.1: Selected intersections.

City	Intersection	Signalization	Crosswalk Marking
Seattle, WA	S1	Signalized	Bar-pair
	S2	Unsignalized	Bar-pair
	S3	Unsignalized	Unmarked
	S4	Unsignalized	Unmarked
	S5	Unsignalized	Unmarked
	S6	Unsignalized	Unmarked
	S7	Signalized	Bar-pair
	S8	Signalized	Bar-pair
	S9	Signalized	Bar-pair
Tampa, FL	T1	Signalized	Continental
	T2	Signalized	Continental
	T3	Signalized	Continental
	T4	Signalized	Standard
	T5	Signalized	Standard
	T6	Signalized	Standard
	T7	Unsignalized	Unmarked
	T8	Unsignalized	Unmarked

3.2.3 Variables

The data used in this study included the forward-view video, face- and hand/dash-view video, vehicle kinematics, demographics of participants, and the crosswalk configuration of the selected intersections (see Table 3.2).

The video files were labeled using video coding software, *Datavyu*. Every traversal was linked back to the SHRP2 demographic database. The vehicle kinematic data and the two parts of video data were synced by matching the timestamp and the unique ID of each traversal. Each instance of secondary task (n = 225) was matched with four instances when drivers were not engaging in any secondary tasks (n = 900) at the same intersection and with the same length of duration. Table 3.4 shows the frequency of non-secondary and secondary tasks.

Figure 3.2: Pictures of a) unmarked, b) standard, c) continental, and d) bar-pair crosswalks from the selected intersection in Seattle and Tampa from Google street view.

- *Forward-view videos* contain video data facing out of the front window of the vehicle. The forward-view video coding started from one block before the selected crosswalk and ended at one block after the crosswalk. Weather conditions, the status of the traffic signals, traffic influence, the presence of pedestrians (including their positions and activities) at the crosswalks were coded from the forward-view videos.

- *Face-view and Hand/dash-view videos* include video data of the driver's face, hands, and vehicle dashboard. The research team coded the drivers' secondary task engagement, the number of hands on the steering wheel (0, 1, or 2), and the presence of front- (yes, no) and back-seat passengers (yes, no, unknown) by matching the start and end time with the forward-view videos. The description of secondary tasks in the work of Zou et al. (2023) was followed for video coding. The nine categories of secondary tasks used in this study are summarized in Table 3.3.

- *Vehicle kinematics.* Measures related to drivers' longitudinal and lateral vehicle control have been shown to reflect the extent of drivers distraction (Ebnali et al., 2016; Osman et al., 2019; Papantoniou et al., 2017; Regan et al., 2008). To identify the similarities of the nine secondary tasks, four driving performance measures were utilized

Table 3.2: Overview of variables obtained from SHRP2 data.

Variables	**Levels**
Forward-view videos	
Weather	Clear, not clear
Traffic signal	Green, red, yellow, unsignalized, no signal perceived
Lead vehicle	Leading vehicle, no leading vehicle
Pedestrian position	Before, after
Pedestrian activity	No pedestrians observed, waiting, walking
Face- and hand/dash-view videos	
Hands-on-wheel	0, 1, 2
Front passenger	Yes, no
Back passenger	Yes, no, unknown
Vehicle kinematics	
Speed	continuous (m/s)
Gas pedal position	continuous (%)
Distance off-center	continuous (m)
Demographics	
Income level	Under $50,000/unknown, $50,000 to $100,000, over $100,000
Working status	Not working/work from home (WFH)/unknown, part-time, full-time
Mileage last year	Under 10,000 miles, 10,000 to 20,000 miles, over 20,000 miles
Business use	Yes, no
Age group	Under 25, 25-70, over 70
Education level	High-school or lower/unknown, college level, post-graduate degree
Crosswalk Configuration	
Crosswalk markings	Unmarked, standard, continental, bar-pair

in this study:

- Vehicle speed: Vehicle speed indicated on speedometer.
- Gas pedal position: Position of the accelerator pedal collected from the vehicle.
- Distance off the lane center: Distance to the left or right of the center of the lane.

The sampling rate of speed was lower than other variables and the research team addressed the missing speed values using the last observation carried forward impu-

Table 3.3: Secondary task description.

Secondary tasks	**Description**
Adjusting in-vehicle devices	Any manipulation towards the center console (adjusting climate control, adjusting radio, inserting CD).
Eating/drinking	Subject drinks from a container or has food placed in their mouth.
Interaction with passengers and/or pets	Subject is talking, listening, reaching to or moving away from a pas -senger or a pet.
Cell phone or tablet operation	Subject press buttons or interacts with a touch screen on a cell phone or a tablet.
Hygiene	Any personal hygiene related behaviors (e.g., combing/brushing/fixing hair, apply make-up, shaving, brushing/flossing teeth, biting nails).
Reaching for object	Subject is reaching for any item (e.g., cell phone, food, cigarette, dropped item, personal body-related items, cloths) without any further manipulation (e.g., cell phone dialing).
Reading	1) Subject vehicle driver is reading material that is in the vehicle, but not a part of the vehicle (e.g., not reading external signs, or center stack display). 2) Subject is holding and reading a cell phone or tablet without pressing any buttons.
Talking and/or listening to a device	1) Subject vehicle driver is talking on a phone or has the phone up to ear as if listening to a phone conversation or waiting for the person they are calling to pick up the phone. 2) Subject vehicle driver is moving lips as if talking or singing, the inter -action is not believed to be with a passenger.
Other external/internal distractions	1) Subject is looking outside the vehicle (eyes off the forward roadway, possibly distracted by construction, pedestrian, previous crash, etc.). 2) Subject vehicle driver glances away from the direction of travel at something inside the subject vehicle but cannot determine a specific glance location.

tation.

- In this study, *demographic data* provided for each SHRP2 participant can be traced back through each traversal obtained. Six variables were considered: income level, working status, age group, education level, mileage last year, and whether they used the study vehicle was for any business purposes. The education level variable encompasses three levels: high school or lower/unknown, college level (including "some education beyond high school but no degree" and college degree), and post-graduate degree (including "some graduate or professional school, but no post-graduate degree" and post-graduate degree).

- *Crosswalk configuration*: The research team identified intersections containing pedestrian facilities using the SHRP2 Insight database for estimates of traversal counts in areas of interest.

Table 3.4: Frequency of secondary task and matched non-secondary task observations.

Task type	**Frequency**	
	Sec-task	Non-sec task
Adjusting in-vehicle devices	5	20
Eating and/or drinking	11	44
External and/or internal distraction	44	176
Interacting with passengers and/or pets	55	220
Cell phone or tablet operation	14	56
Hygiene	44	176
Reaching for object	7	28
Reading	9	36
Talking and/or listening to a device	36	144
Total	225	900

3.2.4 Data Analysis

A cluster analysis was first conducted on four driving performance measures, including vehicle speed, gas pedal position, distance off center, and the number of hands on the steering wheel, to cluster the nine secondary tasks. Classifying driver secondary task behaviors by driving performance data is useful to examine potential differences of secondary task activities in their effects on driving safety. The cluster analysis provided a framework to identify previously unobserved secondary task behavior in a performative manner.

A mixed-effects logit model is then implemented with an outcome variable of the clustered secondary tasks to identify the explanatory factors on distracted driving at intersections. Predictors in the logit model are derived from video data, and additionally include demographic and crosswalk configuration information.

K-means Clustering

K-means clustering is a widely-used unsupervised learning method to partition a data set (MacQueen, 1967). This study applied the average silhouette width (ASW) to find the optimal number of clusters by evaluating the cluster results using two criteria, separation and compactness (Rousseeuw, 1987). The grouping of secondary tasks was based on similarities in the vehicle speed, gas pedal position, the number of hands on the steering wheel, and the distance off the lane center. The outliers were removed and min-max normalization was applied on each variable. After that, the mean values of the four variables in each instance were computed for k-means clustering to group the secondary tasks.

Mixed-effects Logit Model

A mixed-effects logit model was employed to estimate the likelihood of engaging in each secondary task group relative to not performing any secondary task (baseline). The multicollinearity between variables was checked before fitting the logic model. Intersection was used as the random effect, which account for the variation caused by the intersection differences. The multicollinearity of the model was assessed by the Variable Inflation Factors score (O'brien, 2007). High multicollinearity was observed between some pairs of variables. Those variables that remained in the model included front passenger, pedestrian activity, and crosswalk marking. Those excluded included back passenger, pedestrian position, and speed limit.

3.3 Results

3.3.1 Secondary Tasks Grouping

Figure 3.3 shows that $k = 3$ is the optimal number of clusters based on the ASW. A larger ASW indicates smaller within-cluster variations and larger between-cluster variations in the k-means clustering result. Table 3.5 summarises the mean values of the selected variables for k-means clustering. Group 1 contains the secondary tasks of reaching for object, adjusting in-vehicle devices, and other external/internal distractions. Group 2 consists of talking and/or listening, reading, hygiene, cell phone or tablet operation, and eating and/or drinking. Group 3 includes only interactions with passengers and/or pets.

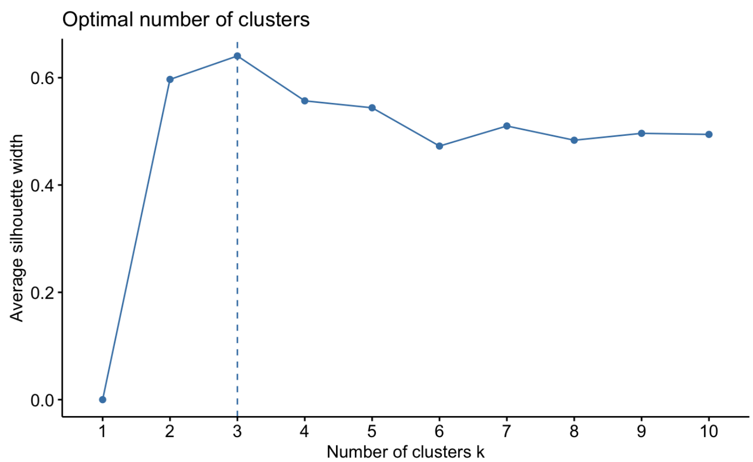


Figure 3.3: Silhouette plot for determining the optimal number of clusters in k-means.

Table 3.5: Summary of the secondary task groups after k-means clustering.

Group	**Secondary task**	**Speed (m/s)**	**Gas pedal position (%)**	**Distance off center (m)**	**Number of Hands-on-wheel**
NA	No secondary task	9.46	10.36	0.35	1.55
1 *Quick visual-manual tasks*	Reaching for objects Adjusting in-vehicle devices Other external/internal distractions	6.12	9.67	0.32	1.19
2 *Continuous visual-manual tasks*	Talking and/or listening to a device Reading Hygiene Cell phone or tablet operation Eating and/or drinking	9.81	10.08	0.39	0.92
3 *Passenger interactions*	Interaction with passenger and/or pets	11.74	9.57	0.45	1.42

Note: Gas pedal position: 0% indicates no pedal depression and 100% indicates pressing the pedal to the floor.

Group 1 tasks included reaching for object, adjusting in-vehicle devices, and other external/internal distractions not identified. Drivers that engaged in Group 1 tasks had the lowest average speed and was able to maintain the closest distance to the lane center, as shown in Table 3.5. This suggests that the drivers had good longitudinal and lateral vehicle control when engaged in these tasks. These tasks were all visual-manual and required glances away from the road. The average duration of these tasks was the shortest compared to other groups (see Table 3.6). Since these tasks include both visual and manual operation, Group 1 will now be identified as *quick visual-manual tasks*.

Group 2 tasks included talking/listening, reading, hygiene, cell phone/tablet operation, and eating/drinking. Drivers engaged in Group 2 tasks had the highest gas pedal position, indicating they used the gas pedal more than the other groups. It also had the smallest averaged number of hands on the steering wheel and the second-largest distance off the lane center, indicating drivers had a low level of lateral control when doing secondary tasks in this group. According to the description in Table 3.3, reading and talking and/or listening to a device include drivers' manipulation of their phones. During the SHRP2 program, there was a lack of voice assistants, and interacting with phones relied on touchscreens and physical buttons. In general, the five secondary tasks in Group 2 were the longest and required both visual and manual engagement. Group 2 will now be referred to as *continuous visual-manual tasks*.

Group 3 only includes interactions with passengers and/or pets. Despite having the highest speed and the largest distance off-center, when drivers engaged in secondary tasks in this group, they maintain their hands on the steering wheel more often compared to other two groups. The average duration of interactions with passengers and/or pets was 38.6 seconds. This group will be named *passenger interactions* for subsequent discussions.

Table 3.6: Average duration of secondary task groups.

Group	**Average duration (s)**
1 *Quick visual-manual tasks*	24.59
2 *Continuous visual-manual tasks*	39.10
3 *Passenger interactions*	38.62

3.3.2 Results of Mixed-effects Logit Model

The final mixed-effects logit model is shown in Table 3.7 (*Quick visual-manual tasks* v.s. baseline), 3.8 (*continuous visual-manual tasks* v.s. baseline), and 3.9 (*passenger interactions* v.s. baseline). The Akaike Information Criterion (AIC) score and Bayesian Information Criterion (BIC) of the mixed-effects logit model are 1429.75 and 1791.59, respectively.

Quick Visual-manual Task vs Baseline

The parameter estimates for this mixed effects logit model *quick visual-manual tasks* compared to no secondary tasks) is shown in Table 3.7. Drivers were 5.2 times more likely to perform *quick visual-manual tasks* while waiting at red lights when compared to driving through green lights. Drivers approaching a continental crosswalk were less likely to engage in these secondary tasks when compared to unmarked crosswalks. The presence of one or more pedestrians waiting at the intersection increased the likelihood of engaging in these tasks by 2.2 folds. The model also showed that part-time workers were less likely to engage in these secondary tasks when compared to full time workers. Drivers with postgraduate degrees were more likely to engage in these secondary tasks when compared to those with up to one college degree. Furthermore, if the vehicle was for business use, the driver was 87.5% less likely to engage in these tasks.

Continuous Visual-manual Tasks vs Baseline

Unlike the *Quick visual-manual tasks*, none of the environmental factors significantly impacted the likelihood of engaging in *continuous visual-manual tasks*. This suggests that these tasks were mainly based on characteristics of the driver (see Table 3.4). More specifically, full-time workers were 3.4 times more likely to engage in *continuous visual-manual tasks* compared to individuals who were either unemployed, WFH, or whose work status was unknown. Senior drivers 70 years and older were 78.7% less likely to engage in these secondary tasks compared to young drivers under 25.

Passenger Interactions v.s. Baseline

Drivers were more likely to interact with passengers when they were seating in the front (44.7 times more likely). They were no significant passenger interactions observed with passengers seating in the back seat. Unlike *quick visual-manual tasks*, driver were more likely to interact with their passengers when approaching transver and continental crosswalks compared with unmarked crosswalks. Drivers with a full-time job were more inclined to interact with passengers than those with the other types of working status. Those whose income were between $50,000 and $100,000 were more likely to interact with passengers compared to those who with lower (under $50,000 or unknown) or higher incomes (over $100,000). Additionally, if the vehicle was for business use, the driver was 87.1% less likely to interact with passengers.

3.4 Discussions

Naturalistic driving data was used to identify factors that contribute to drivers' engagement in various secondary tasks. The nine secondary tasks were grouped using cluster analysis, which revealed three groups: *quick visual-manual tasks*, *continuous visual-manual tasks*, and *passenger interactions*. The results of the mixed-effects logit model indicate that drivers' engagement in *quick visual-manual tasks* was significantly affected by the crosswalk marking designs and the presence of pedestrians. The presence of pedestrians and some demographic factors also played critical roles in the engagement of these secondary tasks. However, the likelihood of engaging in *continuous visual-manual tasks* was only influenced by drivers' demographic factors. Besides, engaging in *passenger interactions* largely depended on the presence of a front passenger, but were also affected by crosswalk marking design and demographic characteristics. This study offers a better understanding of drivers' decision-making in secondary task engagement and provide valuable insights into intervention design.

3.4.1 Groups of Secondary Task

Engaging in *quick visual-manual tasks* requires a diversion of visual attention from the roadway and hands off the steering wheel. According to the multiple resource theory, these secondary tasks interfere with driving because both are visual-manual activities and likely to complete for attentional resources (Wickens, 2008). The cluster analysis shows that drivers had good longitudinal and lateral vehicle control while conducting these *quick visual-manual tasks*. Such adaptions can be seen as compensatory behaviors to reduce the overall demands for attentional resources without compromising driving safety (Kimura et al., 2022; Oviedo-Trespalacios et al., 2016). Previous research has linked these compensatory behaviors to drivers' risk perception in distracted driving situations (Oviedo-Trespalacios et al., 2016). The higher adaptation of speed and lane position observed during *quick visual-manual tasks* suggests that drivers were more aware of the potential risks associated with these tasks compared to other groups. The findings in crosswalk markings suggest that continental crosswalks were effective in reducing the likelihood of engaging in *quick visual-manual tasks*. The pedestrian activities also showed significant impact on the engagement in these tasks. Interestingly, drivers were more likely to engage in *quick visual-manual tasks* when pedestrians were just waiting at an intersection. There were no observed differences between pedestrians walking and no pedestrian presence. The category of "other external/internal distractions" within this group includes the situations that drivers looking at pedestrians other than the forward roadway, which explains why there was a higher likelihood with pedestrians waiting. This finding suggests that drivers might perceive a higher level of uncertainty and allocate more attention on pedestrians when they are waiting by the side, as they could either decide to cross the street or continue waiting.

Drivers exhibited a lower level of lateral control and used gas pedal more when doing secondary tasks in *continuous visual-manual tasks* group, indicating that drivers were less likely to take compensatory behaviors while performing these tasks. The *continuous visual-manual tasks* group includes most routine daily activities, such as personal hygiene and cell phone operation. Although these tasks require both visual and manual engagement and could interfere with the primary task of driving, drivers often view them as less cognitive demanding due to their familiarity with these routine activities (Fuller, 2005; Oviedo-Trespalacios et al., 2016). However, given the deterioration in drivers' vehicle control and the fact that these tasks tend to last longer than others, the *continuous visual-manual tasks* remain hazardous and can lead to severe consequences for pedestrian safety. The likelihood of engagement in these tasks appears to be influenced only by drivers' demographic characteristics. Understanding these relationships can help in the development of targeted interventions to enhance pedestrians safety at intersections. For example, specialized training programs can be developed to increase awareness about the negative effects of driving while engaging in routine daily activities among drivers with a high school education or less and those who are employed full-time.

The engagement of *passenger interactions* largely depends on the presence of the front passenger and the demographic information of the drivers. Leveraging the multiple resource theory, *passenger interactions* involve auditory input and verbal processing, which have less interference with driving compared to the *quick visual-manual* and *continuous visual-manual* groups. Prior studies have indicated that talking and/or listening to passengers are usually considered as low-risk by drivers (Young and Lenné, 2010), which might explain the observation that drivers were even inclined to interact with passengers while approaching standard and continental crosswalks compared to unmarked crosswalks.

3.4.2 Individual Factors

In this study, demographic characteristics were found to influence how drivers engaged in various types of secondary tasks. Some of these factors are closely related to risk perception, which in turn shapes distracted driving behavior (Cox et al., 2023; Jing et al., 2023; Machado-León et al., 2016; Ye et al., 2017). Specifically, older adults exhibit higher risk perception compared to younger drivers (Machado-León et al., 2016; Rhodes and Pivik, 2011). Drivers with higher mileage may develop better risk assessments and are less likely to underestimate the dangers of distraction (Chen et al., 2022). These findings suggest that risk perception mediates the relationship between these demographic factors and secondary task engagement in driving. Beyond risk perception, education and income levels can also influence how drivers view their multitasking abilities. According to reports by Schroeder et al. (2018) and AAA Foundation for Traffic Safety (2023), more educated and higher-income drivers reported more frequent phone use while driving. Moreover, previous studies have suggested that employed drivers are likely to engage in secondary tasks during commute due to their habitual behavior or obligation to handle work-related duties (Costantini et al., 2022; Teodorovicz et al., 2022). These results underscore the importance of investigating the underlying factors behind these demographic characteristics. By integrating demographic profiles with behavioral mediators like risk perception, targeted interventions can be developed to more effectively enhance driving safety.

3.4.3 Limitations

This study's findings are subject to certain limitations. First, the data from the SHRP2 NDS dataset does not fully represent the current US driving population but instead reflects participants from five cities who volunteered for the study. Flannagan et al. (2019) also addressed the limitation that young drivers (16-24 years old) were over-represented in SHRP2 when compared to the actual driver population. Additionally, the dataset was collected between 2010 and 2012. Since that time, there has been more widely adopted in-vehicle systems that encourage drivers to engage in virtual meetings, cell phone or tablet operations. Future research can employ other datasets to eliminate the bias caused by the SHRP2 dataset. Second, the proposed logit model only uncovers the predictors for the engagement in secondary tasks, while the factors that contribute to disengagement from secondary tasks remain unknown. Future research should investigate the predictors of disengagement to gain a more comprehensive understanding of distracted driving. Third, though naturalistic driving data provides us rich driving-related information in real-world conditions, such observational study has limited information outside the vehicles. Further research can consider conducting a controlled study to understand the motivation behind distracted driving by capturing a more comprehensive range of contextual variables.

Table 3.7: Mixed-effects logit model: *quick visual-manual tasks* v.s. Baseline (non-secondary task).

Variables	**Coefficient**	**Odds ratio**	***p***
Front passenger - reference level: no			
Front passenger: yes	0.384	1.468	0.364
Traffic signal - reference level: no signal perceived			
Traffic signal: green	-0.093	0.911	0.894
Traffic signal: red	1.832	6.246	0.006 **
Traffic signal: yellow	0.330	1.391	0.728
Traffic signal: unsignalized	-1.143	0.319	0.110
Marking type - reference level: unmarked			
Marking type: standard	-1.147	0.318	0.198
Marking type: continental	-2.350	0.095	0.019 *
Marking type: bar-pair	-0.577	0.562	0.561
Pedestrian activity - reference level: no pedestrians			
Pedestrian activity: waiting	1.165	3.206	0.034 *
Pedestrian activity: walking	-0.656	0.519	0.561
Working status - reference level: full-time			
Working status: part-time	-1.987	0.137	0.003 **
Working status: not working/WFH/unknown	-0.650	0.522	0.148
Age group - reference level: 25-70			
Age group: under 25	0.923	2.517	0.117
Age group: over 70	-0.048	0.953	0.920
Income - reference level: 50k-100k			
Income: under $50,000/unknown	-0.595	0.552	0.150
Income: over $100,000	-0.874	0.417	0.060 .
Education - reference level: college level			
Education: high school or lower/unknown	0.349	1.418	0.587
Education: advanced	0.947	2.578	0.020 *
Miles last year - reference level: 10k-20k			
Miles last year: under 10,000 miles	-0.132	0.876	0.767
Miles last year: over 20,000 miles	0.065	1.067	0.912
Business use - reference level: no			
Business use: yes	-2.078	0.125	0.007 **

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$

Table 3.8: Mixed-effects logit model: *continuous visual-manual tasks* v.s. Baseline (non-secondary task).

Variables	**Coefficient**	**Odds ratio**	***p***
Front passenger - reference level: no			
Front passenger: yes	-0.532	0.597	0.212
Traffic signal - reference level: no signal perceived			
Traffic signal: green	-0.386	0.680	0.286
Traffic signal: red	0.310	1.363	0.431
Traffic signal: yellow	-0.715	0.489	0.276
Traffic signal: unsignalized	0.097	1.102	0.866
Marking type - reference level: unmarked			
Marking type: standard	-0.032	0.969	0.960
Marking type: continental	-0.293	0.746	0.667
Marking type: bar-pair	-0.167	0.846	0.843
Pedestrian activity - reference level: no pedestrians			
Pedestrian activity: waiting	0.289	1.335	0.488
Pedestrian activity: walking	0.363	1.438	0.447
Working status - reference level: full-time			
Working status: part-time	-0.686	0.504	0.065 .
Working status: not working/WFH/unknown	-1.225	0.294	0.000 ***
Age group - reference level: 25-70			
Age group: under 25	0.567	1.763	0.125
Age group: over 70	-0.959	0.383	0.030 *
Income - reference level: $50,000-$100,000			
Income: under $50,000/unknown	-0.404	0.668	0.174
Income: over $100,000	-0.302	0.739	0.374
Education - reference level: college level			
Education: high school or lower/unknown	0.051	1.052	0.916
Education: advanced	-0.099	0.906	0.762
Miles last year - reference level: 10,000-20,000 miles			
Miles last year: under 10,000 miles	-0.485	0.616	0.110
Miles last year: over 20,000 miles	-1.058	0.347	0.017 *
Business use - reference level: no			
Business use: yes	0.480	1.616	0.299

**** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$*

Table 3.9: Mixed-effects logistic regression model: *passenger interactions* v.s. Baseline (non-secondary task).

Variables	**Coefficient**	**Odds ratio**	***p***
Front passenger - reference level: no			
Front passenger: yes	3.853	47.134	0.000 ***
Traffic signal - reference level: no signal perceived			
Traffic signal: green	-0.354	0.702	0.512
Traffic signal: red	-1.442	0.236	0.077 .
Traffic signal: yellow	-20.589	0.000	0.998
Traffic signal: unsignalized	1.158	3.184	0.218
Marking type - reference level: unmarked			
Marking type: standard	2.025	7.576	0.042 *
Marking type: continental	2.470	11.822	0.021 *
Marking type: bar-pair	1.260	3.525	0.370
Pedestrian activity - reference level: no pedestrians			
Pedestrian activity: waiting	0.854	2.349	0.183
Pedestrian activity: walking	-17.860	0.000	0.998
Working status - reference level: full-time			
Working status: part-time	-1.722	0.179	0.019 *
Working status: not working/WFH/unknown	-1.206	0.299	0.021 *
Age group - reference level: 25-70			
Age group: under 25	0.680	1.974	0.331
Age group: over 70	-0.549	0.578	0.348
Income - reference level: 50k-100k			
Income: under $50,000/unknown	-1.758	0.172	0.000 ***
Income: over $100,000	-2.087	0.124	0.000 ***
Education - reference level: college level			
Education: high school or lower/unknown	1.517	4.559	0.020 *
Education: advanced	-0.275	0.760	0.597
Miles last year - reference level: 10k-20k			
Miles last year: under 10,000 miles	0.281	1.324	0.568
Miles last year: over 20,000 miles	0.261	1.298	0.598
Business use - reference level: no			
Business use: yes	-1.521	0.218	0.113

**** p <0.001, ** p <0.01, * p <0.05, . p <0.1*

Chapter 4

MODELING MULTITASKING PERFORMANCE BY IDENTIFYING FACTORS AND NEUROPHYSIOLOGICAL INDICATORS

In Chapter 4, I explored factors and neurophysiological indicators of multitasking performance through a laboratory study. The Multi-Attribute Task Battery II (MATB-II), an experimental platform that simulates pilots' routine tasks on a flight deck, was employed as the multitasking environment. Sixteen college-aged students were recruited, and their behavioral responses, eye movements, and EEG signals were collected during MATB-II. Chapter 4 revealed the significant impact of task complexity and repetition on multitasking performance. Correlation analysis revealed significant relationships between multitasking errors and fixation count, fixation duration, central alpha-band power, frontal alpha-band power, central theta-band power, and frontal theta-band power. The connection between selected neurophysiological responses and the number of multitasking errors indicates the potential use of portable EEG and eye-tracking devices in real-world scenarios to monitor changes in multitasking performance. Chapter 4 also lays the groundwork for Chapter 5, where the selected factors and indicators are used as contextual and observable variables in the proposed DBN. The results in Chapter 4 and 5 have been published in Li et al. (2025).

4.1 Introduction

4.1.1 Factors Impacting Multitasking Performance

Task complexity and task repetition have been reported to affect multitasking performance, as discussed in Section 2.3.1. One way to describe task complexity is the level of attentional resources required to conduct a particular task (Li and Kim, 2021; Robinson, 2001). Substantial studies have indicated that greater task complexity leads to lower multitasking performance (Liu et al., 2016; Puma et al., 2018; Valéry et al., 2019). According to Valéry et al. (2019), no matter the complexity of the increments in a single task or any combination of tasks in multitasking, a decrease in multitasking performance is expected. Puma et al. (2018) found that as the level of task complexity became greater, achieved by increasing the number of activated sub-tasks in a pilot simulator, there was a decline in the overall performance score. As for task repetition, studies have found that repetition increases the predictability of tasks, allowing individuals to allocate their attentional resources more effectively across tasks and leading to improved multitasking performance (Broeker et al., 2018; Ewolds et al., 2021).

Individual differences among operators are also known to influence multitasking. Age is one of the most commonly discussed factors, with younger adults typically outperforming older adults (Kievit et al., 2014; Reimers and Maylor, 2005; Todorov et al., 2014; Wechsler et al., 2018), as discussed in Section 2.3.2. Todorov et al. (2014) observed lower performance in the older group compared to the younger group on a multitasking test battery that tests executive functioning and spatial ability. Wechsler et al. (2018) examined scenarios combining dual-task and task switching in a driving simulator, where older drivers demonstrated lower driving performance and a higher risk of incidents.

4.1.2 Selecting Factors and Indicators for Multitasking Performance

Feature selection is a fundamental part for model development, as it identifies the most relevant features that contribute to or reflect changes in multitasking performance. Due to the high-dimensional nature of neurophysiological data, feature selection is crucial for reducing dimensionality and improving the modeling efficiency. Furthermore, removing redundant features can help mitigate the potential for over-fitting, particularly when dealing with relatively small sample sizes (Li et al., 2017).

4.2 Methodology

4.2.1 Participants

16 participants (ten women and six men) were recruited with an average of 19.88 years old (SD = 1.26). Each of the participants was a student at the University of Washington. In an online pre-experimental questionnaire, participants were asked about their demographics, eyesight, and prior knowledge of task settings. All of them had normal or corrected-to-

normal eyesight. Each participant performed four experimental trials, and 64 observations were obtained in total.

4.2.2 Apparatus

Participants performed multitasking experiments using a 15-inch Dell laptop in the Human and System Laboratory at the University of Washington. The screen resolution was 1920 x 1080 at a scale of 100% of the screen. The viewing distance between the participants and the laptop screen was controlled at 40 cm. A monitor-mounted Tobii Pro X3-120 eye tracker (Tobii) was used to record participants' eye movements. All the eye movement data were recorded in Tobii Pro Studio (Tobii). I utilized a 32-channel portable EEG device to capture brain signals. The EEG data were recorded using EmotivPro (Emotiv). To minimize movement artifacts in EEG signals caused by physical movement or movement between the scalp and electrodes, throughout the experiment, all participants were required to sit in a chair and avoid making any significant physical movements.

Participants completed the MATB-II developed by NASA, a computer-based flight simulator software program (Santiago-Espada et al., 2011). The MATB-II platform was selected for its standardization and generality given that the MATB-II has been widely used in experimental laboratories to test operators' multitasking performance (Comstock Jr and Arnegard, 1992; Santiago-Espada et al., 2011). During the experiment with MATB-II, the participants were instructed to conduct concurrently the three tasks–resource management, system monitoring, and communication–that mimic operators' typical task sets in flight. Figure 4.1 shows the main user interface of MATB-II. Tracking task was not activated in this study. The simultaneous operations of multiple tasks are MATB-II's key feature, which also represents the nature of most other operational systems that require the ability to handle the demands of multiple tasks concurrently. I controlled the time intervals between the tasks to manipulate task complexity levels in this study. In the NASA-TLX, tasks appear randomly with average time intervals of approximately 6.3 seconds for low-complexity levels and 1.5 seconds for high-complexity levels. MATB-II recorded the timing and the status of each response made by participants during the experiment.

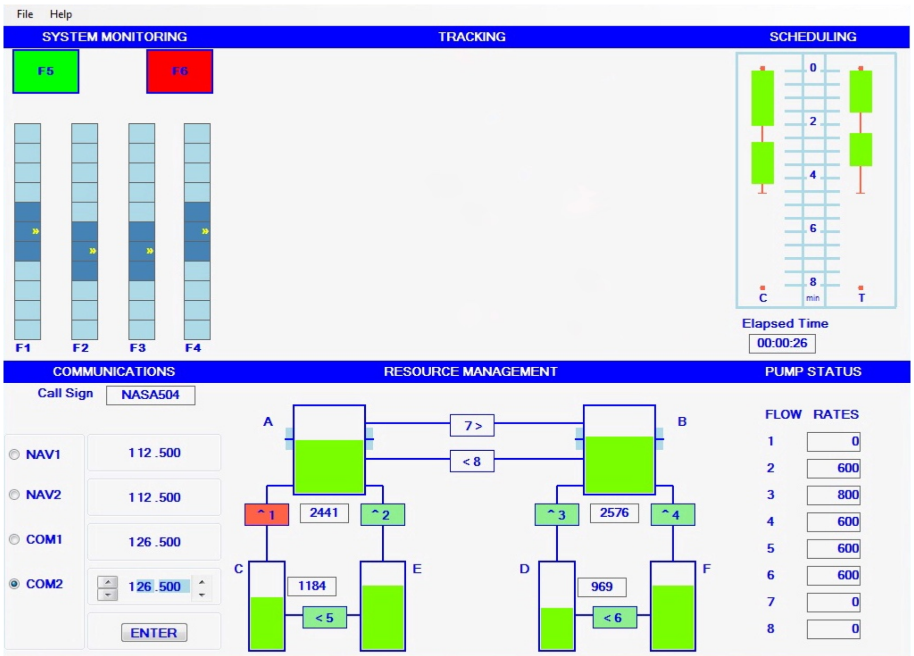


Figure 4.1: The user interface of MATB-II.

4.2.3 Experimental Procedure

We conducted the experiment in the Human and System Laboratory at the University of Washington. Before beginning, each participant was instructed on the procedures and signed a consent form. Then each participant conducted a five-minute practice trial to familiarize themselves with the MATB-II task. After the practice trial, I set up the eye tracker and EEG device. The EEG Ag/Cl electrodes were attached to each participant's scalp with electric gel. According to the previous studies considering frontal (Bohle et al., 2019; Fairclough et al., 2005; Ozdemir et al., 2016; Puma et al., 2018), and central brain regions (Bohle et al., 2019; Fairclough et al., 2005; Ozdemir et al., 2016), 23 electrodes were selected from the corresponding regions based on the International 10–20 system (Sazgar and Young, 2019)(frontal: Fp1, Fp2, F7, F3, Fz, F4, F8; central: FC5, FC1, FC2, FC6, C3, Cz, C4, CP5, CP1, CP2, CP6; parietal: P7, P3, Pz, P4, P8).

The main experiment included two blocks, and each block included both low- and high-complexity trials, resulting in each participant completing a total of four trials. Each trial, either with low or high complexity, lasted eight minutes. I randomized the trial order in each block to avoid the order effect in the experiment. The sequence of the blocks reflected task repetition. Specifically, participants did both low- and high-complexity MATB-II for the first time in the first block, and they repeated both levels of MATB-II in the second block. While participants conducted the MATB-II, their eye movements and brain signals were recorded. During the MATB-II task, participants needed to perform multiple tasks simultaneously: communication (COMM), system monitoring (SYSMON), and resource management (RESMAN). In the COMM task, participants were required to follow radio instructions to select radio stations and adjust the radio frequency accordingly via left-clicking the mouse. The selection of an incorrect radio station or frequency was counted as an error. The SYSMON task included two subtasks. The first subtask required participants to respond when the green light on the interface was turned off by left-clicking the green light or pressing the F5 key and to respond similarly when the red light was turned on by left-clicking it or pressing the F6 key. In the second subtask, participants were asked to monitor four moving scales and to maintain the yellow arrows next to the scales in the center areas. If participants observed the arrows were offset, they needed to press the corresponding keys on the keyboard (F1, F2, F3, and F4 for each scale). I counted it as an error if participants missed or responded incorrectly to an event in SYSMON. When operating the RESMAN task, participants needed to maintain the fuel levels in two target tanks (Tanks A and B) within an acceptable range. Participants could transfer fuel between the target tanks and supplementary tanks (Tanks C, D, E, and F) by controlling the pumps between them. Pumps could be activated or deactivated by left-clicking the mouse or pressing the corresponding keys on the keyboard. If the fuel level in either target tank was below or beyond the range, the label of the tank would turn red, which was considered an error in this study. The entire experiment lasted about 90 minutes.

4.2.4 Analysis

Data Preprocessing

The schematic diagram in Figure 4.2 presents the entire procedure of data analysis in this study. Prior to DBN modeling, I synchronized the variables in the performance and observable layers, i.e., error count, eye movement, and EEG recordings. I then conducted noise deduction and feature selection on eye movement and EEG raw data. Finally, I conducted DBN modeling using the selected variables. I evaluated and validated the modeling result and assessed the capability of the model in multitasking performance compared to other competitive algorithms.

Table 4.1: Types and levels of variables

Type	Variable
Task-related and individual factors	
Discrete	Task complexity (low, high), block (1, 2)
Continuous	Age
Behavioral responses	
Continuous	Error count
Neurophysiological reponses	
Continuous	Fixation count, fixation duration, alpha- and theta-band power at frontal, central, and parietal regions

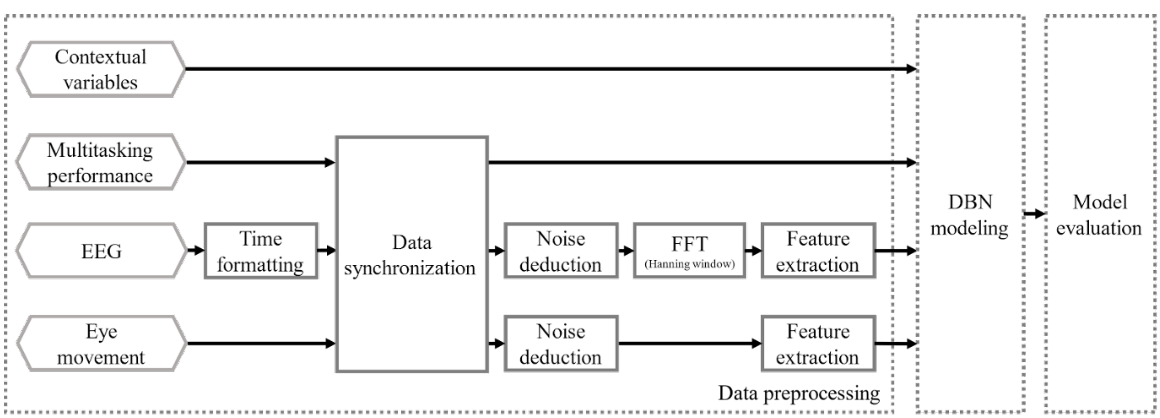


Figure 4.2: Diagram of data processing and modeling

Time Formatting and Data Synchronization

To match the sample rates of datasets recorded by the eye tracker (120 Hz) and EEG (128 Hz), I first converted the epoch time of EEG raw data into the date–time format. I used the start and end of each experimental trial in MATB-II to locate and tailor the eye movement and EEG data to the length of each trial. Then, I down-sampled the EEG and eye movement data different length of time slice, with the shortest time slice as 5 seconds. I considered 5-second time slice because the time interval between tasks in low complexity level was around 6.3 seconds. Time slices shorter than 5 seconds are less likely to include participants' responses to MATB-II.

Noise Deduction and Feature Extraction in Neurophysiological Variables

Before processing the EEG raw data, I checked the contact quality of the 32 electrodes and the scalp of each participant. The Fast Fourier Transform (FFT) based on Welch's method was conducted to decompose the voltage oscillations of EEG into a power spectrum and applied a Hanning function to smooth the amplitude of the discontinuities (Welch, 1967). I utilized the Butterworth band pass filter with a cutoff frequency of 0.5 Hz–60 Hz (Murugappan et al., 2014) and filtered the data based on the frequency ranges of the delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (30–60 Hz) bands. The average alpha-band and theta-band power across the three brain regions were computed. I conducted the EEG data processing and filtering in Python 3 (Van Rossum and Drake, 2009) using the 'Scipy' package (Virtanen et al., 2020). For eye movements, for each time slice, I calculated the average fixation duration and counted the number of fixations.

4.3 Results

4.3.1 Descriptive Statistics

Table 4.2 summarizes the mean and standard deviation of variables for the behavioral and neurophysiological responses by the levels of task complexity and repetition. Participants made more multitasking errors for the high-complexity levels compared to the low-complexity levels, and made more errors in the first blocks compared to the second blocks.

Table 4.2: Descriptive statistics of continuous variables by task complexity and repetition

Variable	Task Complexity		Block	
	Low-complexity	High-complexity	First block	Second block
Performance layer				
Error count	11.00 (9.21)	18.44 (19.48)	18.09 (17.32)	11.34 (13.02)
Observable layer				
Fixation count	1266.16 (141.75)	1302.38 (148.33)	1291.34 (136.28)	1277.19 (155.23)
Fixation duration	311.64 (60.16)	326.09 (60.59)	320.37 (49.86)	317.36 (70.06)
Frontal alpha-band power	19.25 (11.26)	18.94 (9.33)	18.89 (9.44)	19.31 (11.16)
Frontal theta-band power	51.73 (31.10)	52.75 (37.90)	51.78 (31.93)	52.70 (37.21)
Central alpha-band power	6.91 (2.80)	6.65 (2.22)	6.60 (2.34)	6.96 (2.69)
Central theta-band power	12.75 (4.51)	13.12 (5.48)	12.70 (4.80)	13.16 (5.22)

4.3.2 Effects of Task Complexity and Repetition on Multitasking Performance

Non-parametric repeated-measures ANOVA results show that participants made more multitasking errors during the high-difficulty task compared to the low-difficulty task ($F = 103.012$, $p < 0.001$) and during the first block compared to the second block ($F = 15.470$, $p < 0.001$). Figure 4.3 shows the interaction plot of multitasking error as a function of the two factors.

4.3.3 Relationships Between Multitasking Performance and Neurophysiological Responses

Repeated measures correlation tests revealed significant correlations between multitasking performance and the following variables: fixation count ($p < 0.001$), fixation duration ($p < 0.001$), central alpha-band power ($p < 0.01$), frontal alpha-band power ($p < 0.05$), central theta-band power ($p < 0.05$), and frontal theta-band power ($p < 0.01$). Age ($p = 0.953$), alpha- ($p = 0.302$) and theta-band power ($p = 0.336$) in the parietal area did not show significance and were excluded from the further analysis. The continuous variables were

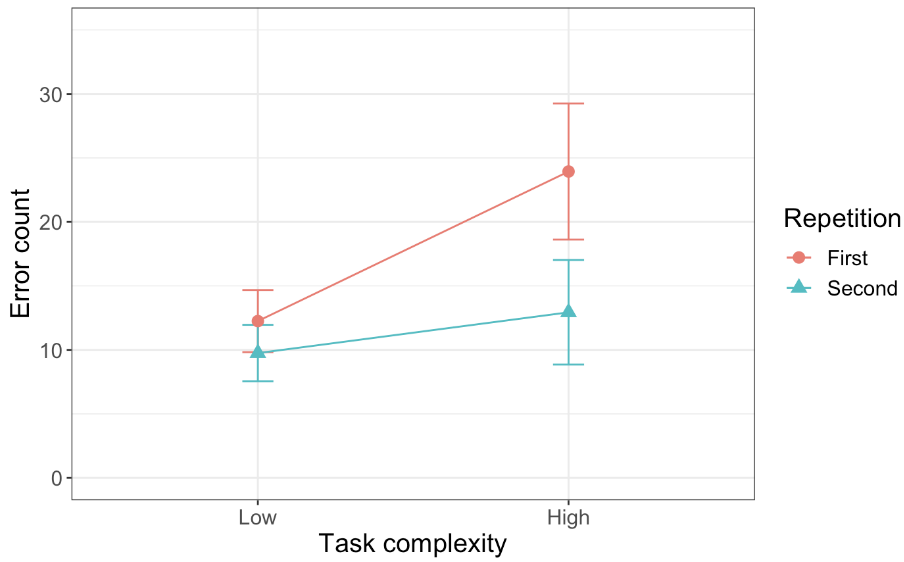


Figure 4.3: Interaction plot for the multitasking error. Lower values indicate better multitasking performance. Error bars represent standard errors.

normalized before the correlation analysis. Table 4.3 shows the correlation test results.

Table 4.3: Repeated-measures correlation test results of continuous variables and error count.

	Age	Fixation count	Fixation duration	Central alpha-band power	Central theta-band power	Frontal alpha-band power	Frontal theta-band power	Parietal alpha-band power	Parietal theta-band power
p	0.953	<0.001***	<0.001***	<0.001***	0.016*	0.037*	0.002**	0.302	0.336
r_{rm}	0.000	-0.099	0.094	-0.044	-0.031	-0.027	-0.040	-0.013	-0.012

4.4 Discussions

In this chapter, I examined the relationships of multitasking performance with various factors and neurophysiological indicators using a software-based multitasking battery, MATB-II. The number of errors made by participants increased with task complexity but decreased with task repetition in MATB-II. Previous studies have not reached a consensus on the relationship between fixation duration and multitasking performance. The correlation between shorter fixations and fewer multitasking errors in this study aligns with findings from Gartenberg et al. (2011) and Hodgetts et al. (2015). Quicker fixations in this study indicated a faster speed of information encoding and processing across all the sub-tasks (Hodgetts et al., 2015). This might ultimately lead to better overall performance. Among EEG band powers, I observed a relationship between increased theta- and alpha-band power in the frontal and central regions with enhanced multitasking performance, as indicated by negative coefficients in the repeated measures correlation analysis. The results in alpha-band power aligns with the notion that suppressed alpha-band activity indicates increasing task demand on attentional resources (Arana et al., 2022; Klimesch, 1999; Puma et al., 2018). However, the finding in theta-band power showed opposite relationship with multitasking performance from previous studies. This observation can be explained by the notion that increased theta-band power may indicate the involvement of greater attentional resources and executive functions, leading to improved multitasking performance (Cavanagh and Frank, 2014; Fairclough and Venables, 2006; Vidulich and Tsang, 2012). There was a lack of significant correlation between age and multitasking performance, although previous studies have shown decline in multitasking performance as age increases (Kievit et al., 2014; Reimers and Maylor, 2005; Todorov et al., 2014; Wechsler et al., 2018). This might due to all the participants were college-aged students at the University of Washington with a relatively narrow age range. Consequently, the non-significant impact of age on multitasking performance could be attributed to the minimal variance within the age group.

This chapter recommends several strategies to enhance multitasking performance in modern workplaces. From the perspective of task design, lowering task frequency, reducing unnecessary information, and providing adequate training can effectively prevent multitasking errors. With the integration of real-time data processing capabilities in recent EEG and eye-tracking devices, this study also demonstrates the potential for using these sensors to monitor operators' multitasking performance, thereby enabling timely intervention to correct errors or mitigate the consequences.

Chapter 5

PREDICTING MULTITASKING PERFORMANCE USING DYNAMIC BAYESIAN NETWORK

Chapter 5 aims to address RQ3 using DBN, a Bayesian-based probabilistic model. Despite the importance of accurately modeling and predicting multitasking performance, most existing multitasking studies have relied on post-hoc measurements that cannot reflect changes in multitasking performance over time. To overcome such limitations, I established a DBN model for the modeling and prediction of multitasking errors using the time-series data generated from the eye-tracker and EEG. The metrics of eye movements and EEG band power were selected from Chapter 4. The effectiveness of the DBN was validated using a comparative analysis with other more commonly used models. This DBN model has the potential to anticipate the occurrence of multitasking errors over time, ultimately improving multitasking strategies in workplace settings.

5.1 Introduction

One of the main challenges of using physiological responses in real-world settings is the low signal-to-noise ratio. Real-world settings, unlike controlled laboratory environments, are often surrounded by uncertain information. A way to mitigate such noisy data is to apply a probabilistic model that identifies the stochastic distribution of variables and their relationships (Murphy et al., 2002). Uncertainties in modeling are covered by the probabilistic distribution of variables. Accordingly, probabilistic models are able to present the effect of uncertainties on the entire system (Ghahramani, 2015).

A DBN is one promising probabilistic modeling technique for capturing the causal relationships among multiple variables over time (Murphy et al., 2002). Using a directed acyclic graph, the DBN includes multiple nodes and arcs. The nodes represent random variables, and the arcs indicate the influential relations (i.e., conditional dependencies) among the variables (Mihajlovic and Petkovic, 2001). The conditional probabilities in DBN are derived based on Bayes' theorem, which calculates posterior probabilities given prior probability and likelihood (Mihajlovic and Petkovic, 2001). DBN represents the time dimension in the form of time slices and collects a series of Bayesian networks that each describes the state of the system in one particular time slice (Li and Ji, 2004). Thus, the movement between the Bayesian networks reflects the temporal change in the system. The temporal relationship in DBN can be explained by the hidden Markov model, meaning that the state of a random variable at a time slice t depends only on its historical state at the previous time slice t-1 and the state of other related variables at the time slice t (Neapolitan and Jiang, 2010).

To build DBN models, some researchers predetermine the variables and their dependencies based on domain knowledge. Doing so simplifies the learning process of DBN structures—i.e., the arcs and nodes—and improves computational efficiency (Reichenberg, 2018; Zeng and Ji, 2010). Domain knowledge may also reduce the chance of false causal relationships between variables when the size of the training dataset is small (Zeng and Ji, 2010).

5.2 Methodology

I proposed a three-layer DBN with the contextual layer as the first layer. This layer included task complexity and the block of the experiment, all of which were treated as discrete variables. In the second layer, I defined the sum of errors in the COMM, SYSMON, and RESMAN tasks as the measure of participants' multitasking performance. In the observable layer, eye movements and EEG band power as treated as continuous data to avoid the loss of information.

I used EEG and eye movements as observable variables to overcome the limitations of relying solely on behavioral performance to continuously predict multitasking performance. I used the eye tracker with a 120 Hz sample rate and EEG with a 128 Hz sample rate, resulting in more than 120 data points recorded per second. Such rich datasets produce reliable predictive models of multitasking performance, which cannot be gathered with behavioral responses. This structure is further supported by previous research that has placed neurophysiological responses in the final observable layer (Liao et al., 2005; Yang et al., 2010).

5.2.1 DBN Modeling

DBN is structured graphical modeling technique that combines multivariate probabilistic systems with Bayes's theorem. Bayes's theorem states that the posterior probability can be computed from the prior probabilities and likelihood, as shown in (5.1).

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \tag{5.1}$$

In this study, I applied a DBN that assumes the state of the variable Z at time t (Z_t) depends only on the state of Z at time $t-1$ (Z_{t-1}) and the parent nodes in the current time slice t. The conditional probability of the variable Z at time t can thus be defined by Equation (5.2), where $Pa^i(Z_t)$ denotes the ith parent nodes of Z_t in the current time slice t.

$$P(Z_t|Z_{t-1}, Pa^1(Z_t), \ldots, Pa^i(Z_t)) = P(Z_t|Z_{t-1}) \prod_{i=1}^{I} P(Z_t|Pa^i(Z_t)) \tag{5.2}$$

According to the chain rule, the resulting joint distribution can be expressed as follows:

$$P(Z_{1:T}) = P(Z_1) \times \prod_{t=2}^{T} \left(P(Z_t|Z_{t-1}) \times \prod_{i=1}^{I} P(Z_t|\mathrm{Pa}^i(Z_t)) \right) \tag{5.3}$$

In DBN, I assume continuous variables follow a normal distribution $N(\mu, \sigma)$. If continuous variable Z has categorical parent nodes, its conditional distribution can be expressed as Equation (5.4), where $Pa(Z)$ represents the parent node of the variable Z where μ_c is the mean and σ_c is the standard derivation (Murphy et al., 2002).

$$P(Pa(Z) = c) = N(z; \mu_c, \sigma_c) \tag{5.4}$$

The conditional distribution of continuous variable Z with continuous parent nodes is shown in Equation (5.5), where W represents the weight matrix (Murphy et al., 2002).

$$P(Pa(Z) = y) = N(z; W \cdot y + \mu, \sigma) \tag{5.5}$$

5.2.2 Time Slices in DBN

In DBN, the spatial relationships and the conditional probabilities among the variables in each time slice are assumed to be static (Bielza and Larranaga, 2014). Changing the length of the time slice in DBN not only affects the relationships among the variables within a time slice but also influences the transition between time slices, which in turn impacts the prediction accuracy of the DBN, as the length of the time slice essentially determines the time scale of the dynamic process (Nodelman et al., 2012). Normally, the length of the time slice is predetermined. In this study, I applied variant lengths of time slices to test the effect of the length on the DBN prediction. 5, 10, 15, 20, 30, 40, and 60 seconds were selected to make sure the data in each experimental trial (480 seconds) could be equally divided.

5.2.3 DBN Validation

This study performed leave-one-out cross-validations to assess the prediction accuracy of DBN and other competitive models to avoid bias caused by using the same participant's data for both training and testing (Wong, 2015). In each iteration of cross-validation, one participant's data was picked as the test set and the remaining 15 participants' data was considered the training set. Mean squared error (MSE) measures the average squared difference between actual values and predictive values (Chicco et al., 2021; Wallach and Goffinet, 1989). Equation (5.6) shows the formulas of MSE, where n is the number of interactions, A_i is the actual value, and P_i is the predicted value. After the 16 iterations of leave-one-out cross-validation, i.e., each participant's data being used as the test set once, the average MSE was computed.

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(A_i - P_i)^2 \tag{5.6}$$

5.3 Results

5.3.1 Structure of the DBN

A three-layered DBN structure is shown in Figure 5.1. The contextual layer includes the block and the task complexity level; the performance layer includes the multitasking performance variable (i.e., the error counts of MATB-II); and the observable layer includes the two features of eye movement and four features of EEG band power. The solid arcs in Figure 5.1 represent the direction of conditional dependencies between variables within a time slice, and the dashed arcs represent the dependencies between variables in two adjacent time slices (Ji et al., 2006). More specifically, the variables in the performance and observable layers in the previous time slice were used as parental nodes to predict the corresponding variables in the current time slice. Since the variables in the contextual layer do not vary over time in the experiment, there are no dashed arcs connecting the contextual variables between time slices.

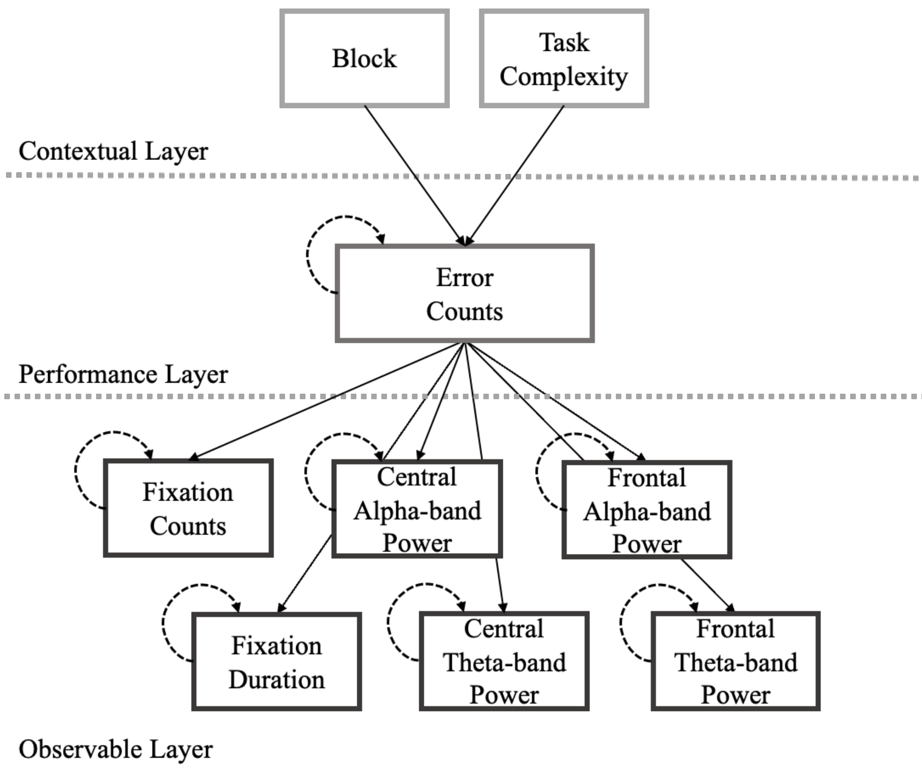


Figure 5.1: Structure of the DBN for predicting multitasking performance.

5.3.2 Prediction Accuracy With Variant Time Slices

Figure 5.2 shows the prediction accuracy of multitasking performance given the different time slice lengths of –5, 10, 15, 20, 30, 40, and 60 seconds. In Figure 5.2, the x-axis shows the time slice length, and the y-axis represents the MSE of the DBN models after cross-validation in terms of each time slice. Smaller prediction errors, as indicated by reduced MSEs, were observed as the length of the time slice decreased from 60 seconds to 5 seconds.

The result revealed that the DBN with the shortest time slice was the most accurate in predicting multitasking performance.

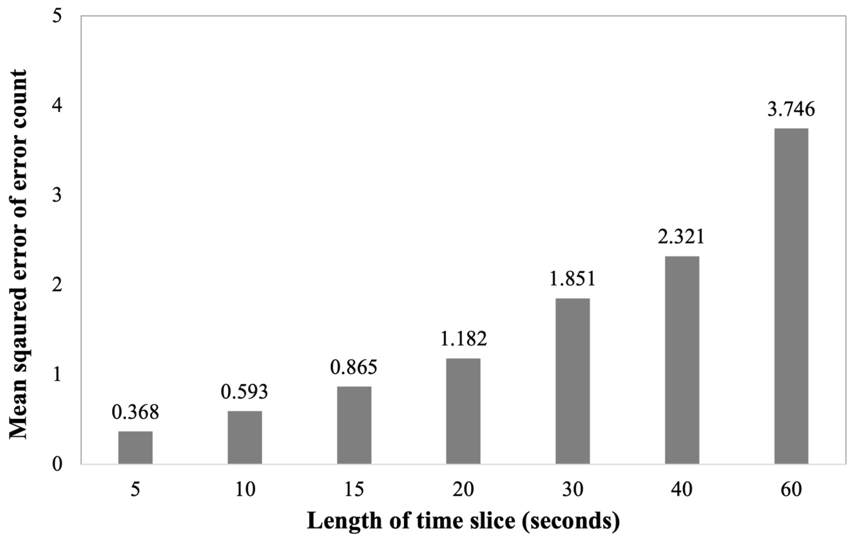


Figure 5.2: Prediction accuracy of DBN using different time slices. The lower mean squared error of error counts represents greater prediction accuracy.

I applied bootstrap resampling to evaluate the robustness of the three-layered DBN model with the current sample size. The bootstrap simulates a large number of datasets by randomly sampling the original dataset (Bland and Altman, 2015). During each of the 100 iterations of the bootstrap (Rosenfeld et al., 2017), a resampled was generated dataset with the same size as the original datasets, fit the DBN model to the resampled data, and calculated the corresponding MSE for each time slice. Table 5.1 shows the MSE of the DBN models after leave-one-out cross-validation and the confidence interval of MSE after bootstrapping in terms of each time slice. The MSE values obtained using leave-one-out cross-validation are within the 95% confidence interval derived from bootstrapping, indicating the robustness of the DBN model given the current sample size.

Table 5.1: Prediction accuracy (MSE) of DBN using leave-one-out cross-validations and the confidence interval of MSE after bootstrapping.

Time slice	**Leave-one-out cross-validation**		**Bootstrap**	
	Mean (SD)	95% CI	Mean (SD)	95% CI
5 seconds	0.368 (0.180)	[0.257, 0.480]	0.362 (0.071)	[0.348, 0.376]
10 seconds	0.593 (0.232)	[0.437, 0.749]	0.601 (0.103)	[0.580, 0.621]
15 seconds	0.865 (0.328)	[0.657, 1.073]	0.854 (0.125)	[0.830, 0.879]
20 seconds	1.182 (0.390)	[0.873, 1.490]	1.216 (0.213)	[1.173, 1.258]
30 seconds	1.851 (0.611)	[1.301, 2.402]	1.902 (0.343)	[1.834, 1.970]
40 seconds	2.321 (0.653)	[1.501, 3.140]	2.331 (0.578)	[2.217, 2.446]
60 seconds	3.746 (1.269)	[2.063, 5.429]	3.663 (1.180)	[3.429, 3.897]

5.3.3 Model Validation by Comparing Prediction Accuracy and BIC Among Different DBN Structures

To evaluate the effectiveness of combining contextual variables and physiological responses in the observable layer of DBN, I compared the prediction accuracy of diverse DBN structures that included specific subsets of the variables. Figure 5.3 presents the proposed DBN (DBN 1) and three variant DBN models that considered EEG power and eye movement only (DBN 2), contextual variables and eye movement only (DBN 3), and contextual variable and EEG power only (DBN 4). The performance of the four DBN models is shown in Table 5.2. The Wilcoxon signed rank test showed significantly smaller MSE in the proposed DBN (DBN 1) compared to DBN 2 ($p < 0.001$), DBN 3 ($p < 0.001$), and DBN 4 ($p < 0.01$). The model comparison among the four DBN models is shown in Table 5.3. Besides, BIC evaluates a model by balancing the goodness of fit and model complexity (Stoica and Selen, 2004), with lower value indicating a better model quality. The proposed DBN achieved the lowest BIC score (-21159.21). In comparison, the BIC scores for DBN 2-4 are -20600.40, -10614.26, and -13001.54, respectively.

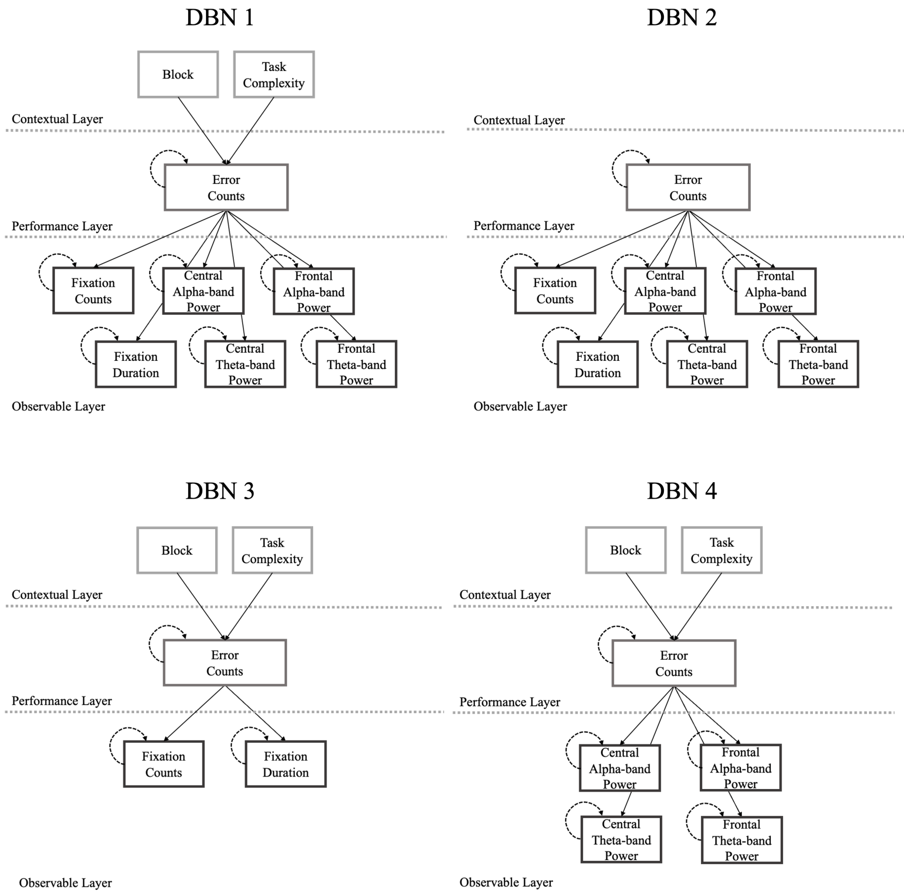


Figure 5.3: The structures of proposed DBN (DBN 1) and its variants (DBN 2, 3, and 4).

Table 5.2: Prediction accuracy of DBNs including contextual, EEG, and eye movements (DBN 1); EEG and eye movements only (DBN 2); contextual and eye movements only (DBN 3); and contextual and EEG only (DBN 4).

Model	**DBN 1**	**DBN 2**	**DBN 3**	**DBN 4**
MSE	1.561	1.606	1.580	1.588
95% CI	[1.038, 2.084]	[1.113, 2.101]	[1.072, 2.087]	[1.057, 2.119]

5.3.4 Prediction Comparison With BN and Other Algorithms

To further evaluate the DBN as an appropriate approach to predict multitasking performance, I compared the DBN's prediction accuracy with that of other popular methods.

Table 5.3: Wilcoxon test results for comparison between DBNs including contextual, EEG, and eye movements (DBN 1); EEG and eye movements only (DBN 2); contextual and eye movements only (DBN 3); and contextual and EEG only (DBN 4).

	Statistics	**DBN 2**	**DBN 3**	**DBN 4**
DBN	*p-value*	<0.001***	<0.001***	0.006**
	W	1862	4703	2226
DBN 2	*p-value*	-	<0.001***	0.002**
	W	-	4436	4232
DBN 3	*p-value*	-	-	0.003**
	W	-	-	2147

Bayesian Network (BN) and several common modeling techniques of linear regression (LR), random forests (RF), and deep neural networks (DNN) were used for comparison. BN can be considered as a single time slice in DBN (Murphy et al., 2002). The BN for comparison has the same structure as DBN, but without the conditional dependencies between the time slices. LR uses linear functions to model relationships among variables (Montgomery et al., 2021). RF can be taken as an aggregate of multiple decision trees, which partitions data into different subgroups based on statistical rules and generates a graphic representation with a tree-like structure (Lan et al., 2020; Song and Ying, 2015). Compared to a single decision tree, a random forest can reduce the variance of the estimated results and achieve higher accuracy in classification and prediction (Breiman, 2001). DNN simulates the functionality of human neurons and is composed of multiple layers of perceptrons. The perceptrons between layers are connected by activation functions and different layers can learn different features, enabling the deep neural network to discover underlying patterns in the dataset (Gurney, 2018).

Table 5.4 shows that the DBN outperformed other competitive models in multitasking performance prediction by showing the smallest average MSE. The results of the Wilcoxon signed rank tests in Table 5.5 showed that DBN had significantly better performance compared to BN, LR, RF, and DNN.

Table 5.4: Prediction accuracy comparison between DBN and other algorithms.

Model	DBN	BN	LR	RF	DNN
MSE	1.561	1.607	2.697	2.562	2.643
95% CI	[1.038, 2.084]	[1.778,3.289]	[1.500, 3.893]	[1.417, 3.707]	[1.232, 4.054]

Table 5.5: Wilcoxon test results for comparison between models.

	Statistics	**BN**	**LR**	**RF**	**DNN**
DBN	*p-value*	<0.001***	<0.001***	<0.001***	<0.001***
	W	668	594	307	449
BN	*p-value*	-	0.553	0.024*	0.073
	W	-	2959	2388	2546
LR	*p-value*	-	-	0.503	0.929
	W	-	-	2933	3195
RF	*p-value*	-	-	-	0.637
	W	-	-	-	3001

5.4 Discussions

I applied a DBN to predict multitasking performance over time using EEG band power and eye movements in this chapter. The DBN model performed better than other DBN models that included only single physiological measurements, indicating the benefit of employing multiple physiological responses in DBN. It has also shown its effectiveness by achieving greater prediction accuracy compared to other modeling approaches. Leveraging the strength of small prediction errors and high computational efficiency, this study suggests the practical uses of the DBN model in improving operators' multitasking performance.

The proposed DBN, which integrates EEG, eye movements, and contextual variables, achieves the highest prediction accuracy and offers a more comprehensive understanding of how multitasking performance relates to task context, cognitive activity, and visual attention. The differences are statistically significant, as confirmed by Wilcoxon tests. Furthermore, the BIC score for the proposed DBN was the lowest among all models, indicating better model fit despite the penalties for relative higher model complexity. Although the absolute difference in prediction error may appear small, the relative difference exceeds 1%. For instance, in the 10-second time slice condition, my DBN achieved a percentage error of 17.6%, compared to 18.9% for DBN 2. These improvements can still be meaningful in safety-critical domains, where small gains in predictive accuracy can accumulate over time and lead to significant operational or safety benefits. The higher prediction accuracy of the DBN model with multisensor measures is consistent with recent research that employs multisensor data in modeling human performance and cognitive states. Multisensor approaches are usually favored because they provide complementary data that single sensors might miss, mitigate the noise susceptibility of single sensors, and ultimately improve prediction accuracy (Al Imran et al., 2024; Brunzini et al., 2024; Iqbal et al., 2024).

The proposed DBN model shows the smallest prediction errors compared to other models, confirming the capability and robustness of the DBN in predicting multitasking performance over time. Compared to a static Bayesian network, the DBN significantly improves multitasking performance prediction by capturing temporal causal relationships among variables, allowing for enhanced prediction over time (Murphy et al., 2002; Yang et al., 2010). The performance of the deep neural network was below my expectations, which may be due to the small size of the training dataset used in this study. Though I might have obtained a better prediction result given a bigger sample size, the training process of the deep neural network would have been more challenging and lengthy (Szandała, 2021), making it less applicable than the DBN to real-world scenarios. The DBN's superior performance compared to other modeling approaches when there is a relatively small sample size may be due to the DBN's unique ability to use the parameter estimations based on conditional probabilities (Reichenberg, 2018) from other models.

The findings in this chapter provide insights into how best to model and predict multitasking performance. The three-layered DBN model provides a framework for predicting multitasking performance using a dynamic probabilistic model. The proposed model has implications beyond the laboratory setting, as error prediction during multitasking is known to be one way to mitigate the consequences of errors (Reason, 1990), especially in high-risk work environments. In particular, if the DBN model detected an increased number of errors in the coming time period, operators could be alerted to the occurrence so that they could react in advance to avoid or minimize the impact of the errors.

Chapter 6

ENHANCING MULTITASKING PERFORMANCE BY AUTOMATION INTERVENTIONS

Chapter 6 examined how to design an automation intervention in a multitasking environment. Partial automation is an effect way to enhance multitasking performance by reducing the overall attention demands of multitasking placed on human operators. Compared to full automation, partial automation requires less technical resources and also reduces operator overreliance on automation by keeping human in the loop. However, there is a critical gap in researchers' understanding of how to design effective partial automation in multitasking, particularly when it comes to determining the types of tasks that should be automated. This study explores the impact of partial automation on multitasking performance under different sensory modalities and task prioritizations. This study utilized the MATB-II as the multitasking environment and employed a 2 × 2 within-subjects design with two independent variables: sensory modality (cross-modality vs. intra-modality) and task priority (equal vs. unequal priority). The individual subtask and overall multitasking performances of 20 participants were measured. Results indicate that partial automation led to greater improvements in both types of performance under cross-modality than under intra-modality. These findings offer guidance for designing automation allocation strategies to enhance multitasking performance.

6.1 Introduction

6.1.1 Sensory Modality and Task Priority

The effectiveness of partial automation depends on an understanding of the factors that influence operator performance during multitasking (Endsley, 2017). Two critical factors—sensory modality and task priority—influence operator multitasking performance and need to be carefully considered when designing automation interventions. Sensory modality refers to the channel through which individuals collect information from environmental stimuli. These stimuli are typically classified based on the sensory receptors they activate (Auvray and Spence, 2008). Visual and auditory modalities are predominant in most safety-critical settings (Hutmacher, 2019). According to multiple resource theory (Wickens, 2002), multitasking interference is greater when tasks use the same sensory modality (e.g., two visual tasks) and lower when tasks use different modalities (e.g., one visual task and one auditory task).

Task priority, another crucial factor, affects multitasking by prompting operators to allocate more attentional resources to high-priority tasks. This prioritization enhances operator performance on these tasks but may lead to decreased performance on tasks of lower priority (Barg-Walkow and Rogers, 2017). In a multitasking scenario involving rerouting, maintenance, and target detection of an unmanned aerial vehicle, participants exhibited higher accuracy and faster response times on the explicitly prioritized task while showing lower accuracy and slower responses on secondary tasks (Liu and Gao, 2024).

6.1.2 Research Question

While research has highlighted the importance of task-related factors in multitasking, key questions remain about the role these factors play in the presence of partial automation. Understanding their influence is essential for designing automation strategies that enhance the system and operator's combined multitasking performance. Specifically, it is unclear which tasks in a multitasking system should be automated given these factors Calhoun (2022). In the following sections of this paper, I refer to the system and operator's combined multitasking performance as "overall multitasking performance". Existing research has only partially addressed this issue. For example, Onnasch (2015) and Karpinsky et al. (2018) investigated the impact of partial automation on operators' attention allocation and overall multitasking performance using the MATB-II. However, both studies preselected the automated task and did not explore other possibilities for task selection. Meanwhile, Dan et al. (2012) conducted a qualitative study examining how automation reliability and task-related factors affect operators' multitasking strategy but did not quantify their impact on overall multitasking performance. Strybel et al. (2016) analyzed how automating different tasks affects overall performance in air traffic control yet did not explain why the automation of certain tasks led to better performance. This study examines how to design partial automation in a multitasking environment, specifically focusing on how to select tasks for automation based on task modality and priority in order to enhance overall multitasking performance.

6.2 Methodology

6.2.1 Participants

Twenty participants (14 women, 6 men) with a mean age of 25.55 years (SD = 4.14) from the University of Washington participated in this study. All participants had normal or corrected-to-normal vision. Individuals with visual or hearing impairments, color blindness, or photosensitive epilepsy, as well as those wearing medical devices that might have been affected by the equipment used in this study, were excluded.

6.2.2 Experimental Design and Procedure

I employed a 2×2 within-subjects design with the two independent variables, sensory modality (i.e., intra- and cross-modality) and task priority (i.e., equal and unequal priority). The level of sensory modality was defined by the combination of two tasks in the multitasking environment. The intra-modality condition involved a visual–visual task combination, while the cross-modality condition featured a visual–auditory task combination. Task priority was determined by the researcher, who before each session specified whether there would be a priority difference between the two tasks (Jansen et al., 2016). In this study, partial automation was implemented by automating one task while keeping the other task manual. The automation's reliability was set at 80%. Only the automated task would be prioritized if there was a priority difference.

Participants who met the inclusion criteria participated in the experiment by visiting the Human and Systems Laboratory at the University of Washington. Upon arrival, participants were asked to consent to participate in the study. Those who consented were provided with verbal instructions related to the experimental procedure. Participants then watched a short video about the experiment design and how to perform the tasks. Next, participants completed practice sessions to familiarize themselves with the keyboard controls. Once participants were ready for the experiment, researchers assisted them in sitting properly in front of the Tobii eye tracker, and conducted the eye-tracker calibration. In the main experiment, each participant completed a total of six trials: two baselines without automation and priority differences, followed by four experimental trials, as shown in Figure 6.1. After each session, participants completed a web-based NASA-TLX questionnaire and took a 1-minute break. After completing all six sessions, participants answered the trust questionnaire. The entire experiment lasted around 90 minutes. Participants were compensated with a $10 Tango card.

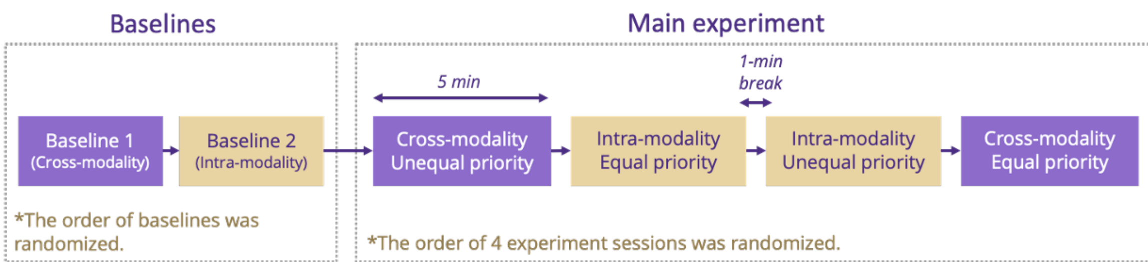


Figure 6.1: Design of experiment.

The dependent variables in this study include the individual performance measures for each subtask and a combined measure representing subtasks' combined performance. The description of each dependent variable can be found in Section 2.3. Table 6.1 lists all the independent and dependent variables.

MATB-II, a computer-based flight simulation software program developed by the National Aeronautics and Space Administration (NASA) (Santiago-Espada et al., 2011), was used as the multitasking environment in this study. I utilized an open-source version of MATB-II, Open-MATB (Cegarra et al., 2020), so that I could create a customized user interface tailored to my research question. Unlike the original MATB-II, this customized setting using Open-MATB showed two subtasks displayed side by side. My study had three subtasks: system monitoring (SYSMON), communication (COMM), and resource management (RESMAN). The left panel featured either the SYSMON or COMM task, depending

Table 6.1: Independent and dependent variables

Category	Sub-category	Variable
Independent variable	-	Task modality (cross-modality, intra-modality)
	-	Task priority (equal, unequal)
Dependent variable	Behavioral response	Automated task: Inverse efficiency score (IES)
		Manual task: Number of abnormal tanks
		Overall performance: Sum of errors
	Subjective response	NASA-TLX
	Physiological response	Number of fixations in automated task
		Number of fixations in manual task
		Number of saccades in automated task
		Number of saccades in manual task
		Dwell time in automated task
		Dwell time in manual task

on the experimental condition, while the right panel displayed the RESMAN task, as illustrated in Figure 6.2. In the four experimental conditions, the task displayed on the left panel (SYSMON or COMM) was automated, while that on the right panel (RESMAN) remained manual.

The intra-modality (visual–visual) condition was created by activating SYSMON and RESMAN at the same time. SYSMON requires participants to respond when they detect abnormalities in any of the four gauges displayed on the interface. When an arrow on a gauge was offset, participants pressed the corresponding key (F1, F2, F3, or F4) on the keyboard to correct it. RESMAN involves maintaining the fuel levels of two target tanks on the airplane within the desired range. As in the SYSMON task, when an arrow on a gauge was offset, participants managed fuel by transferring it between the target tanks (Tank A and Tank B) and supplementary tanks using pumps, which they could activate or deactivate by pressing the corresponding number key on the numpad.

The cross-modality (visual–auditory) condition was created by pairing COMM and RESMAN. The COMM task requires participants to listen to radio messages and adjust the radio

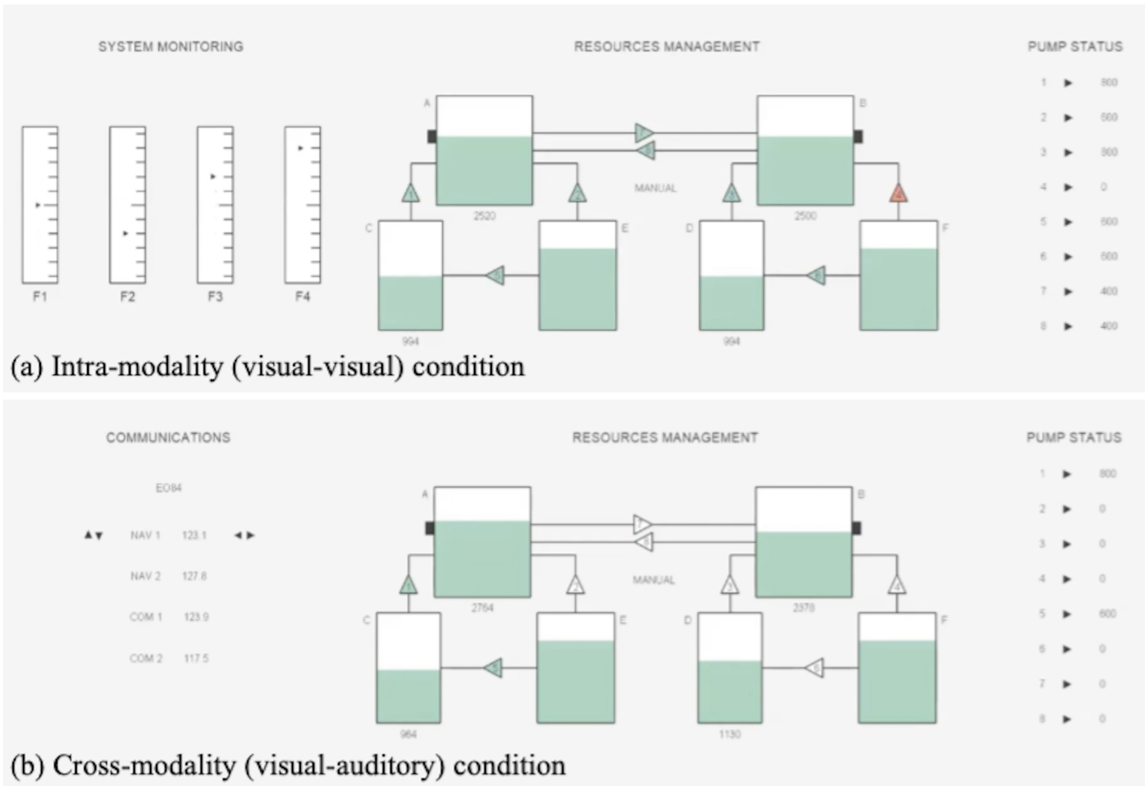


Figure 6.2: User interface of multitasking environment: (a) The intra-modality condition included two visual tasks, system monitoring (SYSMON) and resource management (RESMAN). SYSMON was the automated task, while RESMAN was the manual task; (b) The cross-modality condition included an auditory task, communication (COMM), and a visual task, RESMAN. COMM was the automated task, while RESMAN was the manual task.

channel and frequency based on the messages. The participants used the up and down arrow keys to change the channel and the left and right arrow keys to adjust the frequency. They then confirmed their selection by pressing the return key.

To minimize any potential bias in performance incurred by differences in subtask complexity, I balanced the complexity of SYSMON and COMM by using the Baud rate (Camden et al., 2017), which quantifies the rate of information presentation to the participant, as measured in bits per second (bps). Equation 6.1 shows the Baud rate formula, where B represents the Baud rate of a subtask, N indicates the number of decision alternatives, and ΔT denotes the time interval between two events in the subtask (Liu and Nam, 2018). I controlled the Baud rate for the pair of tasks at 0.25 bps, with SYSMON presenting eight events per minute and COMM presenting three events per minute. The RESMAN task was designed with eight events per minute to balance the complexity of the time-sharing tasks.

$$B = \frac{log_2(N)}{\Delta T} \tag{6.1}$$

6.2.3 Responses

Behavioral Responses

Participants' responses to the subtasks, along with their response times, were recorded by the Open-MATB software. The system's status, such as the fuel levels of Tank A and Tank B, was also logged for analysis. To evaluate overall multitasking performance, I employed a variety of behavioral responses.

- Inverse efficiency score (IES): I used IES, a performance measure that combines response time and the proportion of error responses (Bruyer and Brysbaert, 2011), to generate a single metric by which to evaluate performance on the automated task (SYSMON or COMM). Errors in SYSMON included misses, defined as failures to correct an abnormal gauge, and false alarms, defined as incorrect adjustments to a normal gauge. Errors in COMM were categorized as missing a radio adjustment request, changing a radio channel or frequency without a corresponding request from the radio message, or adjusting the incorrect channel or frequency. Equation 6.2 indicates how to compute IES.

$$\text{IES} = \frac{\text{Response time}}{1 - \text{Error rate}} \tag{6.2}$$

- The number of abnormal tanks: I considered the number of tanks outside the desired fuel level (either above or below), referred to as "abnormal tanks", to evaluate performance on the manual task (RESMAN). The possible values for the number of abnormal tanks were 0, 1, or 2.

- Sum of errors: I used the combined sum of errors from the automated and manual tasks to evaluate overall multitasking performance.

Physiological Responses

I used a Tobii X3-120 eye tracker (Tobii) to record participants' eye movements. The sampling rate was 120 Hz. The distance between the eye tracker and participants was maintained at approximately 500 mm to ensure reliable tracking during the calibration and experiment. The lighting was controlled by blocking out sunlight to ensure consistent experimental settings across participants.

Several metrics were derived from the recorded eye movements to assess participants' attention allocation during multitasking. I defined two AOIs, correspond to the locations of the automated task and the manual task, to analyze task-specific eye movement patterns. Dwell time on each AOI, representing the total fixation time spent within the AOI, was also computed (Becker, 2011). Lastly, I derived the stationary gaze entropy (SGE) from the eye movements, which quantifies the variability of eye movements across the two AOIs and assesses the overall spatial dispersion of fixations (Cui et al., 2024). I used the Shannon's entropy to calculate the gaze entropy as shown in Equation 6.3. n represents the number of AOIs (which is 2 in this study), and p_i is the probability of fixations in the ith AOI.

$$SGE = -\sum_{i=1}^{n} p_i \log_2(p_i) \tag{6.3}$$

Subjective Responses

I used NASA-Task Load Index (NASA-TLX) to measure participants' perceived workload across six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration (Hart, 2006). After rating each dimension, participants performed pairwise comparisons to decide the weights, which reflects the relative contribution of each dimension to the overall workload. The overall workload score was then calculated as a weighted average.

6.2.4 Data Processing and Analysis

I first applied the moving average on the eye movement to smooth the data and reduce the noises. Next, we converted the time formats of the eye movement and MATB-II data by converting them to a local timestamp. The eye movement was downsampled to 1 Hz. Finally, we synchronized the eye movements and behavioral responses, aggregating them by the predetermined time bins to prepare the data for further analysis and modeling.

To normalize the comparisons of behavioral, subjective, and physiological responses across different conditions, I analyzed the relative change in each of the responses from baseline, as shown in Equation 6.4. $X_{\text{experimental}}$ represents the value of a response under the experimental conditions, and X_{baseline} represents its value under the baseline conditions. X_{relative} denotes the relative change in the response, which measures how much the response change in experimental conditions compared to the baseline. A positive relative change indicates an increase in the response, whereas a negative relative change indicates a decrease.

$$X_{\text{relative}} = \frac{(X_{\text{experimental}} - X_{\text{baseline}})}{X_{\text{baseline}}} \tag{6.4}$$

I conducted a non-parametric two-way ANOVA using the aligned rank transform (ART) to examine the effects of sensory modality and task priority on the behavioral responses. ART is a robust method for analyzing factorial designs with non-normal data, allowing for the testing of main effects and interactions (Elkin et al., 2021). The extended ART method adjusts for the repeated measures in my within-subect design by accounting for individual variability across conditions (Beasley, 2002). I used the 'ARTool' package to conduct the analysis in R (R Core Team, 2021).

6.3 Results

6.3.1 Enhancement of Multitasking Performance Through Automation

I first compared the behavioral and subjective responses measured under baseline conditions with those obtained under experimental conditions to determine whether the enhancements in multitasking performance were statistically significant. The Mann–Whitney U test shows significant improvements resulting from automation, indicated by lower IES ($U = 2852.5$, $p < 0.001$), fewer abnormal tanks ($U = 2086$, $p = 0.002$), and a reduced total number of errors ($U = 2338.5$, $p < 0.001$). I also observed a significant difference in NASA-TLX scores between baseline sessions and sessions with automation ($U = 2112$, $p = 0.004$). Table 6.2

summarizes these results, presenting responses in actual values.

Table 6.2: Comparison of behavioral and subjective responses between the baseline and experimental conditions (with automation).

Responses	**Baseline**	**Automation**	***p*-value**
Automated task: IES	2432.506	1237.484	$< 0.001^{***}$
Manual task: Number of abnormal tanks	1.075	0.500	0.002^{**}
Overall performance: Sum of errors	5.109	1.550	$< 0.001^{***}$
NASA-TLX	59.358	48.354	0.004^{**}

6.3.2 Descriptive Statistics

Table 6.3 displays the descriptive statistics of dependent variables for each experimental condition across all participants, including the mean and standard deviation for each variable. Based on the definitions of the three behavioral responses—IES, number of abnormal tanks, and sum of errors—lower values indicate better performance. For NASA-TLX, a higher score indicates a higher self-assessed workload. Responses are presented in actual values. However, in the following sections, only relative changes in experimental conditions compared to baseline will be reported to assess the impact of automation.

6.3.3 Effects of Sensory Modality and Task Priority

Automated Task

Participants showed lower IES on the automated task during cross-modality multitasking compared to intra-modality multitasking ($F = 13.145$, $p < 0.001$). According to the definition of IES, the lower the ratio of the response time to $(1 - \text{Error Rate})$ is, the better the performance of the automated task. The negative values represent improvement in performance compared to baseline. There was no significant effect observed from priority or the interaction of modality and priority. Figure 6.3 depicts the interaction plot for IES across sensory modalities and task priority levels.

Table 6.3: Descriptive statistics of dependent variables (mean and SD) in experimental conditions (with automation). Responses are shown in actual values.

	Task Priority		**Task Modality**	
Dependent Variable	**Equal**	**Unequal**	**Cross-modality**	**Intra-modality**
Automated task: IES	1258.349 (553.037)	1216.619 (553.713)	1031.025 (641.142)	1443.944 (338.882)
Manual task: Number of abnormal tanks	0.400 (0.744)	0.600 (1.057)	0.350 (0.700)	0.650 (1.075)
Overall performance: Sum of errors	1.475 (1.552)	1.625 (1.720)	1.375 (1.444)	1.725 (1.797)
NASA-TLX	48.142 (19.488)	48.567 (17.875)	47.092 (19.445)	49.617 (17.833)

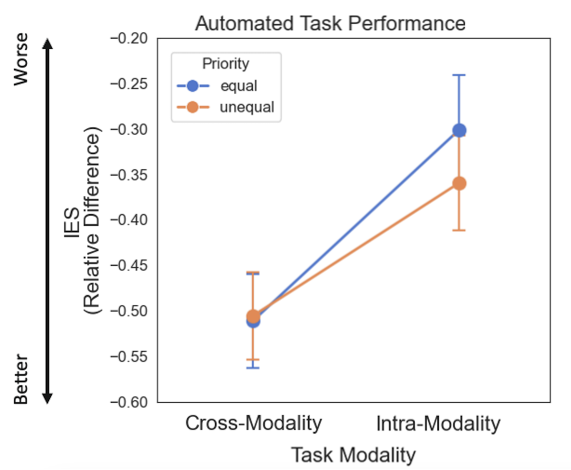


Figure 6.3: Interaction plot for the inverse efficiency score (IES) in the automated task. The y-axis shows the relative difference in IES from baseline. Lower (more negative) values indicate better performance. Error bars represent standard errors.

Manual Task

I observed significantly fewer abnormal tanks during cross-modality multitasking compared to intra-modality multitasking ($F = 9.737$, $p < 0.01$). This suggests better performance on the manual task. There was no significant effect observed from priority or the interaction of modality and priority. Figure 6.4 shows the interaction plot for the number of abnormal tanks across sensory modalities and task priority levels across sensory modalities and task priority levels.

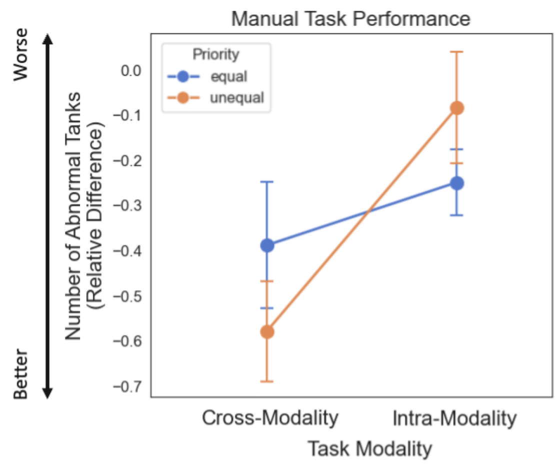


Figure 6.4: The interaction plot for the number of abnormal tanks in the manual task. The y-axis represents the relative change of the number of abnormal tanks from baseline. Lower (more negative) values indicate better performance. Error bars denote standard errors.

Combined Multitasking Performance

With IES and the number of abnormal tanks, participants showed fewer sum of errors under the cross-modality condition than under the intra-modality condition, indicating a better overall multitasking performance under cross-modality ($F = 4.274$, $p < 0.05$). They also showed a marginal effect of task priority on sum of errors, with smaller sums of errors under the equal-priority condition compared to unequal-priority condition ($F = 2.982$, $p = 0.090$). There was no significant interaction effect of modality and priority. Figure 6.5 shows the interaction plot for the combined performance of the automated and manual tasks.

NASA-TLX

There were no significant differences in NASA-TLX scores across sensory modalities or priority levels, suggesting that participants' perceived workload did not vary among the experimental conditions with automation aid.

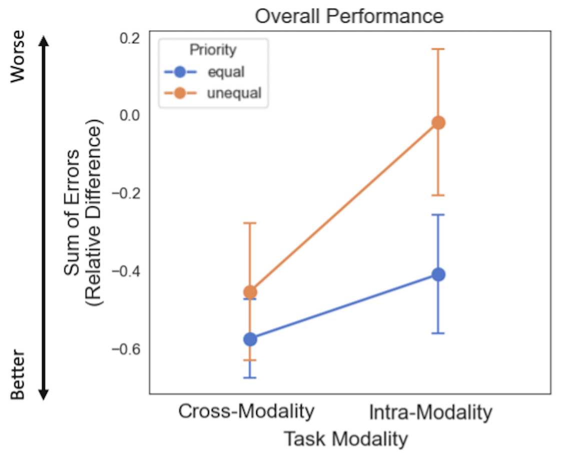


Figure 6.5: The interaction plot for the sum of errors from the automated and manual tasks. The y-axis represents the relative change of the sum of errors from baseline. Lower values indicate better performance. Error bars denote standard errors.

6.4 Discussions

This study examined how sensory modality and task priority influence the effectiveness of partial automation in enhancing overall multitasking performance. These results showed that cross-modality multitasking benefited more than intra-modality multitasking from the incorporation of partial automation, which highlights the importance of tailoring partial automation to account for sensory modalities in multitasking. No significant differences were observed between the equal- and unequal-priority conditions.

6.4.1 Sensory Modality

One explanation for the difference between sensory modalities is the impact of mapping auditory information onto spatially distributed visual elements on MATB-II during cross-modality multitasking. This process likely led to additional spatial transformation demands and increased cognitive effort, resulting in reduced performance Stock et al. (2017); Tsang et al. (2021). In this study, participants had to map channel and frequency information provided in radio messages to visual panels using the MATB-II interface when performing an auditory task (COMM). To assess whether this could explain my results, I conducted a Wilcoxon signed-rank test comparing the cross-modality and intra-modality baselines.

When there was no automation, participants in the intra-modality baseline exhibited better overall multitasking performance compared to those in the cross-modality baseline, as indicated by the former group's fewer sum of errors ($W = 41.5$, $p < 0.05$). Given that task complexity between SYSMON and COMM was balanced, these findings suggest that the introduction of partial automation helped reduce the cognitive effort of performing spatial mapping in cross-modality conditions, leading to larger performance improvements.

Eye movement differences provide additional insights into how participants' allocate their visual attention during multitasking when automation was involved. In cross-modality conditions, participants exhibited longer dwell times in the manual task compared to intra-modality conditions ($F = 39.770$, $p < 0.001$). Longer dwell times are indicative of increased focus on a given area Fichtel et al. (2019), suggesting that participants were able to allocate more visual attention to the manual task in cross-modality scenarios, potentially leading to improved performance on the manual task compared to intra-modality conditions.

6.4.2 Task Priority

To understand the reasons for the insignificant effect of task priority on multitasking performance, I investigated participants' eye movement differences by task priority conditions. When the automated task was assigned a higher priority, participants exhibited longer dwell times on the automated task and shorter dwell times on the manual task than they did under the equal-priority conditions ($F = 16.466$, $p < 0.001$; $F = 16.083$, $p < 0.001$, respectively). This finding suggests that, despite being automated, participants still allocated more visual attention to the prioritized task. While automation is intended to reduce attentional demands on the prioritized task, which usually receives greater attention when competing with other tasks, this advantage may have been offset by increased monitoring of the automated task when it was prioritized.

6.4.3 Correlation Between Eye Movements and Performance Metrics

The repeated measures correlation analysis revealed key eye movement metrics that reflect both subtask performance and overall multitasking performance. A longer average fixation duration, indicating higher cognitive effort for visual information processing, was associated with a higher IES ($r = 0.429$, $p < 0.001$), reflecting lower performance on the automated task. The shorter dwell time in the manual task and the higher SGE were correlated with a higher number of abnormal tanks ($r = -0.268$, $p < 0.05$ and $r = 0.266$, $p < 0.05$, respectively), indicating lower manual task performance. Higher SGE suggests a more dispersed fixation pattern and less efficient visual attention allocation to task-relevant elements. Regarding overall performance, a longer average fixation duration was also associated with greater sum of errors ($r = 0.351$, $p < 0.01$). Additionally, longer dwell time in the automated task, shorter dwell time in the manual task, and higher SGE were correlated with lower overall performance ($r = 0.351$, $p < 0.01$; $r = -0.327$, $p < 0.01$; and $r = 0.330$, $p < 0.01$, respectively). The heatmap in Figure 6.6 shows the results of the correlation analysis.

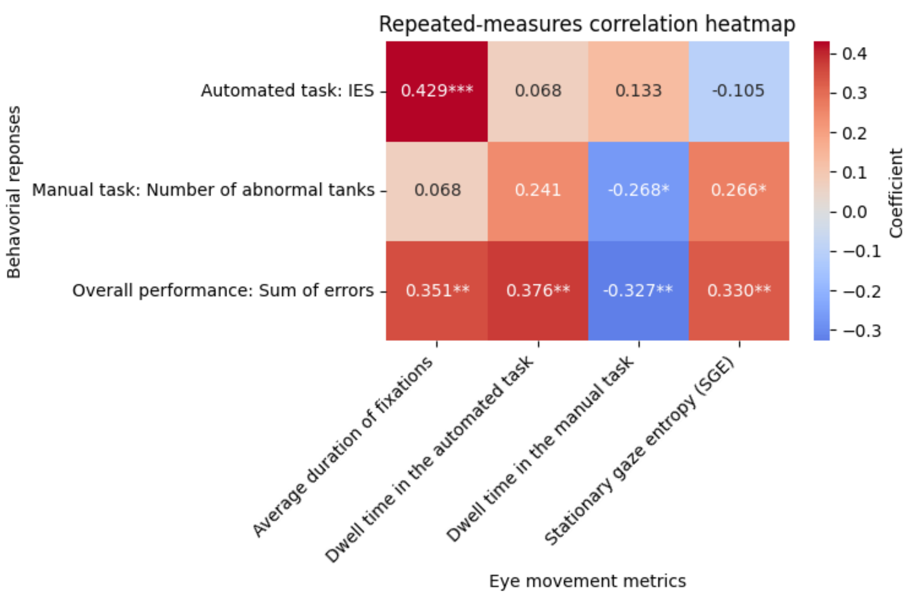


Figure 6.6: Heatmap for correlations between subtask performance, overall multitasking performance, and eye movement metrics. All variables are presented in relative difference format, and the values on the heatmap represent correlation coefficients.

6.4.4 Limitation and Future Research

This study has limitations that require further research. First, I focused on automation with 80% reliability, a value higher than the 70% used by other studies that have shown the efficacy of partial automation in enhancing performance (Onnasch, 2015). Future research could explore various levels of automation reliability to gain a more comprehensive understanding of how to design partial automation. Second, this study examined two levels of task priority, equal priority and unequal priority. The automated tasks, SYSMON and COMM, are both signal-detection tasks that are time-critical Pontiggia et al. (2024). Considering the actual functionality of SYSMON and COMM in flight operations, responding to audio messages from air traffic control and detecting abnormal altitude gauges, are more safety-critical tasks, making them important for maintaining overall system functionality. In contrast, the manual task (RESMAN) involves planning and resource allocation Pontiggia et al. (2024), which is less time-critical and allows for more temporal flexibility. This difference is the primary reason I did not include a condition in which the manual task was prioritized. However, broadening the applicability of these findings would require exploring scenarios in which the manual task is prioritized while partial automation is active. Third, while MATB-II simulates the routine tasks of pilots during flights, exploring the effects of partial automation in real-world multitasking scenarios would provide valuable validation and enhance the generalizability of my findings beyond the simulator-based setting.

Chapter 7

GENERAL DISCUSSIONS AND CONCLUSIONS

7.1 Objectives

The increasing prevalence of multitasking in real-world work settings is driven by the growing complexity of responsibilities and the increasing volume of tasks that operators must process simultaneously (Mark et al., 2005). In high-risk environments such as aviation, driving, and healthcare, operators are required to manage multiple streams of information, coordinate competing task demands, and make accurate decisions under pressure. Performance decrements, such as errors or delayed response times in these settings, can lead to life-threatening consequences, making it imperative to develop data-driven solutions that improve multitasking performance to ensure safety and efficiency. This dissertation developed an integrated framework for multitasking performance enhancement by modeling and predicting multitasking performance and investigating the design of intervention. Table 7.1 summarizes the key findings of this dissertation by research question and chapter.

Effectively modeling and predicting multitasking performance is crucial for mitigating the negative impacts of performance decrements. This dissertation examines multitasking behavior in both a real-world distracted driving scenario and controlled experimental settings, demonstrating how factors such as task complexity, repetition, priority, and sensory modality influence performance. I also identified several time-series neurophysiological metrics from eye movements and EEG signals as reliable indicators to reflect performance changes. Building on these insights, a Bayesian-based dynamic probabilistic model was employed to capture temporal dependencies among these variables, enabling continuous prediction of multitasking performance.

This dissertation also goes one step further by investigating how to design effective intervention support in order to deliver beneficial and timely mitigation once the predictive model detects performance decrements in multitasking. With the rapid development of automation technology, this dissertation also incorporated intervention into the scope and examined how partial automation enhance the performance based on the multitasking environment.

The rest of this chapter concludes by summarizing the main findings and contributions of this dissertation, and discusses the potential directions for future studies.

Table 7.1: Summary of research findings.

Chapter	RQ	Summary of findings
3, 4	RQ1	Contextual factors (crosswalk markings, traffic light, and presence of pedestrians; task complexity and repetition) and individual factors (age group, etc.) influence multitasking performance in field and the laboratory studies.
4	RQ2	Neurophysiological responses, including eye movements (fixation duration and frequency) and brain signals (central and frontal theta-band power and alpha-band power) serve as indicators of changes in multitasking performance.
5	RQ3	A DBN integrating contextual factors and neurophysiological responses provides a robust framework for predicting multitasking performance over time.
6	RQ4	The effectiveness of automation in enhancing multitasking performance is influenced by contextual factors (sensory modality).

7.2 Review of Findings

7.2.1 Factors

Understanding the factors that influence multitasking performance is the first and essential step for modeling multitasking performance. A well-informed model that accounts for key contributors to variations in multitasking performance can provide accurate estimations, enabling timely alerts, and proactive mitigation of performance decrements. The studies collectively identify a range of factors that influence multitasking performance. I investigated the factors that influence distracted driving behavior using naturalistic driving data. Using k-means clustering and multinomial regression, the analysis identifies two broad categories of factors that shape drivers' engagement in secondary tasks and their subsequent impact on driving performance. The first category includes factors related to the driving task, such as traffic signals, the presence of pedestrians, and crosswalk marking types, which influence drivers' situational awareness and decision-making. The second category includes demographic factors, including education level, employment status, age group, driving mileage, and income level, which reflect individual differences in distracting driving. In the controlled study using MATB-II, task-related factors, including task complexity and repetition, as lower complexity and more repetition can result in better multitasking performance. Together, these findings provide insights into how task-related and individual factors influence distracted driving behavior and its consequences on driving performance.

7.2.2 Neurophysiological Indicators

Neurophysiological signals offer distinct advantages over traditional behavioral metrics for assessing multitasking performance. While behavioral measurements provide only post-hoc evaluations and often fail to capture moment-to-moment fluctuations, time-series neurophysiological signals can reflect real-time changes in cognitive and attentional states during multitasking. This dissertation identifies key neurophysiological indicators from eye movements and EEG signals by integrating domain knowledge, which considers their relevance to multitasking, with data-driven methods, specifically repeated measures correlation analysis. Eye movement features, including fixation count and fixation duration, were selected based on their established relationship with visual attention allocation. Similarly, theta- and alpha-band power in the central and frontal regions of EEG were chosen due to their associations with cognitive processes and task-related cognitive demands. These neurophysiological signals demonstrated significant correlations with multitasking errors, validating their potential for tracking performance fluctuations over time. Identifying these physiological indicators lays the groundwork for developing a probabilistic modeling approach to predict multitasking performance.

7.2.3 Probabilistic Modeling

Probabilistic modeling is a powerful method for estimating multitasking performance over time, as it can effectively handle small sample sizes, account for uncertainty, deal with missing data, and provide interpretable representations of variable dependencies. These advantages make probabilistic modeling particularly suitable for multitasking research. In this dissertation, I developed a DBN to predict multitasking performance over time. The model follows a well-validated three-layer structure, incorporating factors in the contextual layer and indicators in the observable layer, which are identified in the previous parts of this dissertation. The model was validated through comparative analysis with other machine learning models, demonstrating its feasibility in estimating multitasking errors over time. Additionally, I evaluated the predictive performance of DBN models incorporating both eye-tracking and EEG signals versus those relying on a single sensor type. The results align with recent trends in human factors research, showing that a multi-sensor approach enhances prediction accuracy compared to single-sensor models.

7.2.4 Interventions

In this dissertation, I introduced partial automation as an intervention to enhance multitasking performance, as it can reduce attentional demands while addressing the limitations of full automation. This section expands on insights from earlier chapters by integrating task-related factors into the automation design discussion and utilizing neurophysiological indicators, specifically eye movements, to gain an understanding of how automation influences human operators' behavior in multitasking environments. This study specifically examines how sensory modality and task priority influence the effectiveness of partial automation. Using non-parametric ANOVAs, the findings reveal that partial automation leads to greater improvements in both subtask and overall multitasking performance when applied under cross-modal conditions compared to intra-modal conditions. This suggests that automation should be strategically allocated between tasks that engage different sensory systems to maximize its benefits.

7.3 Contributions

7.3.1 Scientific Contributions

This dissertation makes several scientific contributions to the fields of human performance modeling and human-automation interaction. First, it introduces a dynamic, data-driven framework for accurately modeling multitasking performance, moving beyond static and post-hoc methods toward an interpretable probabilistic model. Future researchers can easily adjust this model to account for context-specific factors in different task settings and incorporate various physiological sensors to refine its predictions.

Moreover, this framework is not only applicable to multitasking research but also extends to a broad range of domains. One key strength is its flexibility, which allows variables to be added, removed, or reconfigured based on domain-specific theories or empirical findings, enabling adaptation to varied contexts. Beyond this flexibility, the DBN approach also captures temporal relationships among variables, making it well-suited for domains where outcomes fluctuate over time or task demands change. By integrating diverse data sources, including physiological sensors (capturing latent cognitive and physiological states), contextual factors (reflecting task demands), and individual differences, this framework facilitates comprehensive modeling of various outcomes, such as vigilance, fatigue, and cognitive workload. As sensor technologies and data acquisition methods continue to advance, this approach will remain relevant for domains that require adaptive, customized, and time-sensitive modeling.

Finally, this dissertation enhances empirical understanding of automation design in multitasking environments, an area that remains underexplored in current research. It emphasizes the need to study human-automation interactions within the context of specific tasks, which is frequently underrepresented in broader studies. These findings suggest that future research could benefit from this approach that integrates both general principles and context-specific insights. By leveraging insights from task-related factors and neurophysiological indicators, this work contributes a data-driven framework to the design of automation systems in multitasking environment.

7.3.2 Practical Implications

The findings of this dissertation provide actionable insights that can be applied to various domains. First, the proposed probabilistic model incorporating factors and neurophysiological indicators has implications for developing decision support systems, as error prediction is a known strategy for mitigating the consequences of errors in high-risk work environments (Reason, 1990). For example, in the aviation industry, my sensor-based predictive model for multitasking performance could be integrated into pilots' and air traffic controllers' monitoring displays, providing personalized proactive interventions, such as timely performance feedback or alerts for potential performance declines.

Second, my model could be applied to training program design and candidate selection, as maintaining multitasking ability is a key requirement in many modern workplaces. Organizations could use the model to design adaptive training programs that can adjust task design (e.g., task complexity levels) based on trainees' real-time performance, ensuring progressive skill development for roles that require intensive multitasking, such as clinicians. Similarly, the model could be integrated into candidate selection to assess and estimate individuals' multitasking performance over time, helping to identify candidates who are best suited for roles that face complex and dynamic multitasking situations, such as those in military settings.

Third, this dissertation establishes the groundwork for designing context-specific adaptive automation system to enhance operators' multitasking performance. This adaptive automation is particularly beneficial in domains with structured and multitasking-intensive settings, such as driving and aviation, ensuring that automation is deployed at the right moment and on the right task. Understanding how to allocate automation based on task-related factors allows for a more strategic and efficient deployment of automation support. Moreover, by integrating real-time neurophysiological monitoring to detect performance decrements, the system can dynamically trigger context-specific automation that are aligned with the operators' needs and task demands.

7.4 Limitations and Future Rresearch

This dissertation has several limitations offering directions for future research. First, the probabilistic model was developed based on controlled lab studies, validating this framework in another distracted driving study would strengthen the applicability and generalizability of my approach. Second, while demographic factors did not show significant effects in the lab studies and were not included in the DBN model, both the field study in this dissertation and previous research indicate that individual differences can also influence multitasking performance. Besides, the proposed DBN model does not incorporate subtask performance metrics or neurophysiological indicators (e.g., AOI-based eye movement metrics). Future research could enhance the model's generalizability by integrating individual factors, subtask-level performance data and indicators, providing a more detailed and customized prediction models of multitasking performance. Third, the effectiveness of automation was examined in a controlled multitasking setting, which may not fully capture the complexity of real-world multitasking contexts. Future research should investigate how automation can enhance multitasking performance in more dynamic and less well-defined environments, including real-world settings. For instance, future studies could explore how automation allocation strategies should evolve when tasks change over time and user priorities fluctuate, which is very likely to happen in our daily work. These efforts will support the development of more generalized and robust automation interventions to improve multitasking performance. Finally, integrating automation effects into the probabilistic modeling framework would further enhance this dissertation's contributions. A key direction for future research is to extend the DBN model by incorporating automation as an intervention variable. To lay the groundwork for this, I have conducted preliminary research by using repeated-measures correlations between eye movement metrics and both subtask and overall performance (see Figure 6.6). Future studies could build on these findings by integrating significant eye movement metrics as indicators to achieve continuous estimation of multitasking performance with the intervention of automation. This integration would ultimately inform the design of sensor-based adaptive automation systems.

REFERENCES

AAA Foundation for Traffic Safety (2023). 2022 traffic safety culture index (technical report). Technical report, AAA Foundation for Traffic Safety, Washington, D.C. Retrieved from AAA Foundation for Traffic Safety.

Agresti, A. (2012). *Categorical data analysis*, volume 792. John Wiley & Sons.

Ahmad, A., Darmoul, S., Dabwan, A., Alkahtani, M., and Samman, S. (2016). Human error in multitasking environments. In *Proceedings of the International Conference on Industrial Engineering and Operations Management*, pages 1265–1272.

Akhtar, M. J. and Utne, I. B. (2014). Human fatigue's effect on the risk of maritime groundings–a bayesian network modeling approach. *Safety science*, 62:427–440.

Al Imran, M. A., Nasirzadeh, F., and Karmakar, C. (2024). Designing a practical fatigue detection system: A review on recent developments and challenges. *Journal of Safety Research*, 90:100–114.

Arana, L., Melcón, M., Kessel, D., Hoyos, S., Albert, J., Carretié, L., and Capilla, A. (2022). Suppression of alpha-band power underlies exogenous attention to emotional distractors. *Psychophysiology*, 59(9):e14051.

Arbel, Y., Feeley, E., and He, X. (2020). The effect of feedback on attention allocation in category learning: An eye tracking study. *Frontiers in Psychology*, 11:559334.

Aricò, P., Borghini, G., Di Flumeri, G., Bonelli, S., Golfetti, A., Graziani, I., Pozzi, S., Imbert, J.-P., Granger, G., Benhacene, R., et al. (2017). Human factors and neurophysiological metrics in air traffic control: a critical review. *IEEE reviews in biomedical engineering*, 10:250–263.

Auvray, M. and Spence, C. (2008). The multisensory perception of flavor. *Consciousness and cognition*, 17(3):1016–1031.

Avril, E., Valéry, B., Navarro, J., Wioland, L., and Cegarra, J. (2021). Effect of imperfect information and action automation on attentional allocation. *International Journal* *of Human–Computer Interaction*, 37(11):1063–1073.

Barg-Walkow, L. H. and Rogers, W. A. (2017). Modeling task scheduling in complex healthcare environments: Identifying relevant factors. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 61, pages 772–775. SAGE Publications Sage CA: Los Angeles, CA.

Barthelmes, V. M., Heo, Y., Fabi, V., and Corgnati, S. P. (2017). Exploration of the bayesian network framework for modeling window control behaviour. *Building and Environment*, 126:318–330.

Beanland, V., Fitzharris, M., Young, K. L., and Lenné, M. G. (2013). Driver inattention and driver distraction in serious casualty crashes: Data from the australian national crash in-depth study. *Accident Analysis & Prevention*, 54:99–107.

Beasley, T. M. (2002). Multivariate aligned rank test for interactions in multiple group repeated measures designs. *Multivariate Behavioral Research*, 37(2):197–226.

Becker, L., Kaltenegger, H. C., Nowak, D., Rohleder, N., and Weigl, M. (2023). Differences in stress system (re-) activity between single and dual-or multitasking in healthy adults: a systematic review and meta-analysis. *Health psychology review*, 17(1):78–103.

Becker, L., Kaltenegger, H. C., Nowak, D., Weigl, M., and Rohleder, N. (2022). Physiological stress in response to multitasking and work interruptions: Study protocol. *Plos one*, 17(2):e0263785.

Becker, S. I. (2011). Determinants of dwell time in visual search: similarity or perceptual difficulty? *PloS one*, 6(3):e17740.

Bernhardt, K. A., Poltavski, D., Petros, T., Ferraro, F. R., Jorgenson, T., Carlson, C., Drechsel, P., and Iseminger, C. (2019). The effects of dynamic workload and experience on commercially available eeg cognitive state metrics in a high-fidelity air traffic control environment. *Applied ergonomics*, 77:83–91.

Besson, P., Bourdin, C., Bringoux, L., Dousset, E., Maïano, C., Marqueste, T., Mestre, D. R., Gaetan, S., Baudry, J.-P., and Vercher, J.-L. (2013). Effectiveness of physiological and psychological features to estimate helicopter pilots' workload: A bayesian network approach. *IEEE Transactions on Intelligent Transportation Systems*, 14(4):1872–

1881.

Bian, Y., Liang, K., Zhao, X., Li, H., and Yang, L. (2020). Evaluating the effectiveness of new-designed crosswalk markings at intersections in china considering vehicle-pedestrian interaction. *Accident Analysis & Prevention*, 139:105498.

Bielza, C. and Larranaga, P. (2014). Bayesian networks in neuroscience: a survey. *Frontiers in Computational Neuroscience*, 8:131.

Black, S. C., Bender, A. D., Whitney, S. J., Loft, S., and Visser, T. A. (2022). The effect of multi-tasking training on performance, situation awareness, and workload in simulated air traffic control. *Applied cognitive psychology*, 36(4):874–890.

Bland, J. M. and Altman, D. G. (2015). Statistics notes: bootstrap resampling methods. *BMJ*, 350.

Bohle, H., Rimpel, J., Schauenburg, G., Gebel, A., Stelzel, C., Heinzel, S., Rapp, M., and Granacher, U. (2019). Behavioral and neural correlates of cognitive-motor interference during multitasking in young and old adults. *Neural plasticity*, 2019.

Bommer, S. C. and Fendley, M. (2018). A theoretical framework for evaluating mental workload resources in human systems design for manufacturing operations. *International Journal of Industrial Ergonomics*, 63:7–17.

Bonnefond, M. and Jensen, O. (2024). Resisting distraction: On the role of alpha oscillations in gain modulation and resource allocation.

Borys, M. and Plechawska-Wójcik, M. (2017). Eye-tracking metrics in perception and visual attention research. *EJMT*, 3:11–23.

Brams, S., Ziv, G., Levin, O., Spitz, J., Wagemans, J., Williams, A. M., and Helsen, W. F. (2019). The relationship between gaze behavior, expertise, and performance: A systematic review. *Psychological bulletin*, 145(10):980.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Brishtel, I., Ishimaru, S., Augereau, O., Kise, K., and Dengel, A. (2018). Assessing cognitive workload on printed and electronic media using eye-tracker and eda wristband. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion*, pages 1–2.

Broadbent, D. E. (1982). Task combination and selective intake of information. *Acta* *psychologica*, 50(3):253–290.

Broccia, G., Milazzo, P., and Ölveczky, P. C. (2019). Formal modeling and analysis of safety-critical human multitasking. *Innovations in Systems and Software Engineering*, 15:169–190.

Broeker, L., Liepelt, R., Poljac, E., Künzell, S., Ewolds, H., de Oliveira, R. F., and Raab, M. (2018). Multitasking as a choice: A perspective. *Psychological research*, 82(1):12–23.

Brunzini, A., Grandi, F., Peruzzini, M., and Pellicciari, M. (2024). An integrated methodology for the assessment of stress and mental workload applied on virtual training. *International Journal of Computer Integrated Manufacturing*, 37(9):1088–1106.

Bruyer, R. and Brysbaert, M. (2011). Combining speed and accuracy in cognitive psychology: Is the inverse efficiency score (ies) a better dependent variable than the mean reaction time (rt) and the percentage of errors (pe)? *Psychologica Belgica*, 51(1):5–13.

Caldani, S., Razuk, M., Septier, M., Barela, J. A., Delorme, R., Acquaviva, E., and Bucci, M. P. (2019). The effect of dual task on attentional performance in children with adhd. *Frontiers in integrative neuroscience*, 12:67.

Calhoun, G. (2022). Adaptable (not adaptive) automation: Forefront of human–automation teaming. *Human factors*, 64(2):269–277.

Camden, A., Nickels, M., Fendley, M., and Phillips, C. A. (2017). A case for information theory-based modelling of human multitasking performance. *Theoretical issues in ergonomics science*, 18(3):266–278.

Campbell, K. L. (2012). The shrp2 naturalistic driving study: Addressing driver performance and behavior in traffic safety. *Tr News*, (282).

Carsten, O. and Martens, M. H. (2019). How can humans understand their automated cars? hmi principles, problems and solutions. *Cognition, Technology & Work*, 21(1):3–20.

Castet, E. and Crossland, M. (2012). Quantifying eye stability during a fixation task: a review of definitions and methods. *Seeing and Perceiving*, 25(5):449–469.

Cavanagh, J. F. and Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in cognitive sciences*, 18(8):414–421.

Cegarra, J., Valéry, B., Avril, E., Calmettes, C., and Navarro, J. (2020). Openmatb: A multi-attribute task battery promoting task customization, software extensibility and experiment replicability. *Behavior research methods*, 52:1980–1990.

Chen, L., Chen, C., and Ewing, R. (2012). The relative effectiveness of pedestrian safety countermeasures at urban intersections—lessons from a new york city experience. In *Transportation Research Board (TRB) 91st Annual Meeting, Washington, DC*.

Chen, Y., Liu, X., Xu, J., and Liu, H. (2022). Underestimated risk perception characteristics of drivers based on extended theory of planned behavior. *International journal of environmental research and public health*, 19(5):2744.

Chérif, L., Wood, V., Marois, A., Labonté, K., and Vachon, F. (2018). Multitasking in the military: Cognitive consequences and potential solutions. *Applied cognitive psychology*, 32(4):429–439.

Chicco, D., Warrens, M. J., and Jurman, G. (2021). The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *PeerJ Computer Science*, 7:e623.

Choudhary, P., Pawar, N. M., Velaga, N. R., and Pawar, D. S. (2020). Overall performance impairment and crash risk due to distracted driving: A comprehensive analysis using structural equation modelling. *Transportation research part F: traffic psychology and behaviour*, 74:120–138.

Choudhary, P. and Velaga, N. R. (2017a). Analysis of vehicle-based lateral performance measures during distracted driving due to phone use. *Transportation research part F: traffic psychology and behaviour*, 44:120–133.

Choudhary, P. and Velaga, N. R. (2017b). Modelling driver distraction effects due to mobile phone use on reaction time. *Transportation Research Part C: Emerging Technologies*, 77:351–365.

Clark, T., David, Y., Baretich, M., Bauld, T., Dickey, D., Gieras, I., et al. (2006). Impact of clinical alarms on patient safety. *ACCE Healthcare Technology Foundation*, pages 1–20.

Collet, C., Petit, C., Priez, A., and Dittmar, A. (2005). Stroop color–word test, arousal, electrodermal activity and performance in a critical driving situation. *Biological psychology*, 69(2):195–203.

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

Comstock Jr, J. R. and Arnegard, R. J. (1992). The multi-attribute task battery for human operator workload and strategic behavior research. Technical Report NAS 1.15: 104174, NASA.

Costantini, A., Ceschi, A., and Oviedo-Trespalacios, O. (2022). Eyes on the road, hands upon the wheel? reciprocal dynamics between smartphone use while driving and job crafting. *Transportation research part F: traffic psychology and behaviour*, 89:129–142.

Courage, M. L., Bakhtiar, A., Fitzpatrick, C., Kenny, S., and Brandeau, K. (2015). Growing up multitasking: The costs and benefits for cognitive development. *Developmental Review*, 35:5–41.

Cox, A. E., Cicchino, J. B., Reagan, I. J., and Zuby, D. S. (2023). Prevalence of distracted driving by driver characteristics in the united states. *Journal of safety research*, 86:346–356.

Craig, C. M., Morris, N. L., and Hong, Y. (2019a). A case study on the impact of crosswalk markings on driver yielding to pedestrians. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 63, pages 1399–1403. SAGE Publications Sage CA: Los Angeles, CA.

Craig, C. M., Morris, N. L., Van Houten, R., and Mayou, D. (2019b). Pedestrian safety and driver yielding near public transit stops. *Transportation research record*, 2673(1):514–523.

Crews, D. E. and Russ, M. J. (2020). The impact of individual differences on multitasking ability. *International Journal of Productivity and Performance Management*, 69(6):1301–1319.

Cui, Z., Sato, T., Jackson, A., Jayarathna, S., Itoh, M., and Yamani, Y. (2024). Gaze transition entropy as a measure of attention allocation in a dynamic workspace involving automation. *Scientific Reports*, 14(1):23405.

Dan, C. S., Cullen, R. H., Rogers, W. A., and Fisk, A. D. (2012). Exploring strategy use in a multiple-task environment: Effects of automation reliability and task properties. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 56, pages 2123–2127. SAGE Publications Sage CA: Los Angeles, CA.

De Jong, R. (1995). The role of preparation in overlapping-task performance. *The Quarterly journal of experimental psychology*, 48(1):2–25.

De Winter, J. C., Gosling, S. D., and Potter, J. (2016). Comparing the pearson and spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychological methods*, 21(3):273.

Dehais, F., Lafont, A., Roy, R., and Fairclough, S. (2020). A neuroergonomics approach to mental workload, engagement and human performance. *Frontiers in neuroscience*, 14:268.

Delbridge, K. A. (2000). *Individual differences in multi-tasking ability: Exploring a nomological network*. PhD thesis, Michigan State University.

Do, T.-T. N., Chuang, C.-H., Hsiao, S.-J., Lin, C.-T., and Wang, Y.-K. (2019). Neural comodulation of independent brain processes related to multitasking. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(6):1160–1169.

Dogan, E., Steg, L., and Delhomme, P. (2011). The influence of multiple goals on driving behavior: The case of safety, time saving, and fuel saving. *Accident Analysis & Prevention*, 43(5):1635–1643.

Domeyer, J. E., Seaman, S., Angell, L., Lee, J., Reimer, B., Zhang, C., and Donmez, B. (2016). SHRP2 NEST database: Exploring conditions of secondary task engagement in naturalistic trip data. In *Adjunct Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '16 Adjunct, pages 185–190, New York, NY, USA. Association for Computing Machinery.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., et al. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1):27–46.

Douglas, H. E., Raban, M. Z., Walter, S. R., and Westbrook, J. I. (2017). Improving our understanding of multi-tasking in healthcare: Drawing together the cognitive psychology and healthcare literature. *Applied ergonomics*, 59:45–55.

Dux, P. E., Tombu, M. N., Harrison, S., Rogers, B. P., Tong, F., and Marois, R. (2009). Training improves multitasking performance by increasing the speed of information processing in human prefrontal cortex. *Neuron*, 63(1):127–138.

Ebnali, M., Ahmadnezhad, P., Shateri, A., Mazloumi, A., Heidari, M. E., and Nazeri, A. R. (2016). The effects of cognitively demanding dual-task driving condition on elderly people's driving performance; real driving monitoring. *Accident Analysis & Prevention*, 94:198–206.

Eckstein, M. K., Guerra-Carrillo, B., Singley, A. T. M., and Bunge, S. A. (2017). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental cognitive neuroscience*, 25:69–91.

Elkin, L. A., Kay, M., Higgins, J. J., and Wobbrock, J. O. (2021). An aligned rank transform procedure for multifactor contrast tests. In *The 34th annual ACM symposium on user interface software and technology*, pages 754–768.

Endsley, M. R. (2017). From here to autonomy: lessons learned from human–automation research. *Human factors*, 59(1):5–27.

Eom, H., Lee, D., Han, S., Hariyani, Y. S., Lim, Y., Sohn, I., Park, K., and Park, C. (2020). End-to-end deep learning architecture for continuous blood pressure estimation using attention mechanism. *Sensors*, 20(8):2338.

Evans, L. H., Herron, J. E., and Wilding, E. L. (2015). Direct real-time neural evidence for task-set inertia. *Psychological science*, 26(3):284–290.

Ewolds, H., Broeker, L., De Oliveira, R. F., Raab, M., and Künzell, S. (2021). Ways to improve multitasking: Effects of predictability after single-and dual-task training. *Journal of Cognition*, 4(1).

Fairclough, S. H. and Venables, L. (2006). Prediction of subjective states from psychophysiology: A multivariate approach. *Biological psychology*, 71(1):100–110.

Fairclough, S. H., Venables, L., and Tattersall, A. (2005). The influence of task demand and learning on the psychophysiological response. *International Journal of Psychophysiology*, 56(2):171–184.

Ferraro, J. C. and Mouloua, M. (2021). Effects of automation reliability on error detection and attention to auditory stimuli in a multi-tasking environment. *Applied Ergonomics*, 91:103303.

Ferris, T. K. and Sarter, N. (2011). Continuously informing vibrotactile displays in support of attention management and multitasking in anesthesiology. *Human Factors*, 53(6):600–611.

Fichtel, E., Lau, N., Park, J., Henrickson Parker, S., Ponnala, S., Fitzgibbons, S., and Safford, S. D. (2019). Eye tracking in surgical education: gaze-based dynamic area of interest can discriminate adverse events and expertise. *Surgical endoscopy*, 33:2249–2256.

Fintor, E., Stephan, D. N., and Koch, I. (2018). Emerging features of modality mappings in task switching: Modality compatibility requires variability at the level of both stimulus and response modality. *Psychological Research*, 82(1):121–133.

Fischer, R. and Plessow, F. (2015). Efficient multitasking: Parallel versus serial processing of multiple tasks. *Frontiers in psychology*, 6:1366.

Fitzpatrick, K., Chrysler, S. T., Iragavarapu, V., Park, E. S., et al. (2010). Crosswalk marking field visibility study. Technical report, United States. Federal Highway Administration. Office of Safety Research and ....

Flannagan, C. A., Baykas, P. B., Leslie, A., Kovaceva, J., Thomson, R., et al. (2019). Analysis of shrp2 data to understand normal and abnormal driving behavior in work zones. Technical report, United States. Federal Highway Administration.

Foroughi, C. K., Devlin, S., Pak, R., Brown, N. L., Sibley, C., and Coyne, J. T. (2023). Near-perfect automation: Investigating performance, trust, and visual attention allocation. *Human factors*, 65(4):546–561.

Fosch-Villaronga, E., Khanna, P., Drukarch, H., and Custers, B. (2023). The role of humans in surgery automation: Exploring the influence of automation on human–robot interaction and responsibility in surgery innovation. *International Journal of Social Robotics*, 15(3):563–580.

Foxe, J. J. and Snyder, A. C. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Frontiers in psychology*, 2:154.

Freiberger, J. J., Derrick, B., Natoli, M. J., Akushevich, I., Schinazi, E. A., Parker, C., Stolp, B. W., Bennett, P. B., Vann, R. D., Dunworth, S. A., et al. (2016). Assessment of the interaction of hyperbaric n2, co2, and o2 on psychomotor performance in divers. *Journal of applied physiology*, 121(4):953–964.

Friedgen, E., Koch, I., and Stephan, D. N. (2021). Modality compatibility in task switching depends on processing codes and task demands. *Psychological Research*, 85(6):2346–2363.

Fuller, R. (2005). Towards a general theory of driver behaviour. *Accident analysis & prevention*, 37(3):461–472.

Gartenberg, D., McCurry, M., and Trafton, G. (2011). Situation awareness reacquisition in a supervisory control task. In *proceedings of the human factors and ergonomics society annual meeting*, volume 55, pages 355–359. SAGE Publications Sage CA: Los Angeles, CA.

Gevins, A., Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., Du, R., and Rush, G. (1998). Monitoring working memory load during computer-based tasks with eeg pattern recognition methods. *Human factors*, 40(1):79–91.

Ghahramani, Z. (2013). Bayesian non-parametrics and the probabilistic approach to modelling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1984):20110553.

Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. *Nature*, 521(7553):452–459.

Gitelman, V., Carmel, R., Pesahov, F., and Hakkert, S. (2017). An examination of the influence of crosswalk marking removal on pedestrian safety as reflected in road user behaviours. *Transportation research part F: traffic psychology and behaviour*, 46:342–355.

Griffiths, S. W., Brockmark, S., Höjesjö, J., and Johnsson, J. (2004). Coping with divided attention: the advantage of familiarity. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(1540):695–699.

Guastello, A. D., Guastello, S. J., and Guastello, D. D. (2014). Personality trait theory and multitasking performance: Implications for ergonomic design. *Theoretical Issues in Ergonomics Science*, 15(5):432–450.

Guo, H. and Boyle, L. N. (2022). Driving behavior at midblock crosswalks with rectangular rapid flashing beacons: hidden markov model approach using naturalistic data. *Accident Analysis & Prevention*, 165:106406.

Gurney, K. (2018). *An introduction to neural networks*. CRC Press.

Gutzwiller, R. S. and Sitzman, D. M. (2017). Examining task priority effects in multi-task management. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 61, pages 762–766. SAGE Publications Sage CA: Los Angeles, CA.

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

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

Hassani, S., Kelly, E. H., Smith, J., Thorpe, S., Sozzer, F. H., Atchley, P., Sullivan, E., Larson, D., and Vogel, L. C. (2017). Preventing distracted driving among college students: Addressing smartphone use. *Accident Analysis & Prevention*, 99:297–305.

Hellier, E., Naweed, A., Walker, G., Husband, P., and Edworthy, J. (2011). The influence of auditory feedback on speed choice, violations and comfort in a driving simulation game. *Transportation research part F: traffic psychology and behaviour*, 14(6):591–599.

Henelius, A., Korpela, J., and Huotilainen, M. (2011). Individualising eeg frequency bands for sleep deprivation studies. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6083–6086. IEEE.

Heng, K. W. (2014). Teaching and evaluating multitasking ability in emergency medicine residents-what is the best practice? *International journal of emergency medicine*, 7(1):41.

Himi, S. A., Bühner, M., Schwaighofer, M., Klapetek, A., and Hilbert, S. (2019). Multitasking behavior and its related constructs: Executive functions, working memory capacity, relational integration, and divided attention. *Cognition*, 189:275–298.

Himi, S. A., Volberg, G., Bühner, M., and Hilbert, S. (2023). Individual differences in everyday multitasking behavior and its relation to cognition and personality. *Psychological Research*, 87(3):655–685.

Hodgetts, H. M., Tremblay, S., Vallières, B. R., and Vachon, F. (2015). Decision support and vulnerability to interruption in a dynamic multitasking environment. *International*

Journal of Human-Computer Studies, 79:106–117.

Hogervorst, M. A., Brouwer, A. M., and Van Erp, J. B. (2014). Combining and comparing eeg, peripheral physiology and eye-related measures for the assessment of mental workload. *Frontiers in Neuroscience*, 8:322.

Hopkin, V. D. (2017). *Human factors in air traffic control*. CRC Press.

Horrey, W. J., Lesch, M. F., and Garabet, A. (2008). Assessing the awareness of performance decrements in distracted drivers. *Accident Analysis & Prevention*, 40(2):675–682.

Hu, W. and Cicchino, J. B. (2018). An examination of the increases in pedestrian motor-vehicle crash fatalities during 2009–2016. *Journal of safety research*, 67:37–44.

Huestegge, L. (2011). The role of saccades in multitasking: towards an output-related view of eye movements. *Psychological Research*, 75(6):452–465.

Hunt, A. R. and Kingstone, A. (2004). Multisensory executive functioning. *Brain and Cognition*, 55(2):325–327.

Hunter, S. W., Divine, A., Frengopoulos, C., and Montero Odasso, M. (2018). A framework for secondary cognitive and motor tasks in dual-task gait testing in people with mild cognitive impairment. *BMC geriatrics*, 18:1–7.

Hutmacher, F. (2019). Why is there so much more research on vision than on any other sensory modality? *Frontiers in psychology*, 10:481030.

Iani, C. and Wickens, C. D. (2007). Factors affecting task management in aviation. *Human factors*, 49(1):16–24.

Ikehara, C. S. and Crosby, M. E. (2005). Assessing cognitive load with physiological sensors. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, page 295a. IEEE.

Iqbal, M. U., Srinivasan, B., and Srinivasan, R. (2024). Multi-class classification of control room operators' cognitive workload using the fusion of eye-tracking and electroencephalography. *Computers & Chemical Engineering*, 181:108526.

Jackson, A. F. and Bolger, D. J. (2014). The neurophysiological bases of eeg and eeg measurement: A review for the rest of us. *Psychophysiology*, 51(11):1061–1071.

Jansen, R. J., van Egmond, R., and de Ridder, H. (2016). Task prioritization in dual-tasking: Instructions versus preferences. *PLoS One*, 11(7):e0158511.

Ji, Q., Lan, P., and Looney, C. (2006). A probabilistic framework for modeling and real-time monitoring human fatigue. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 36(5):862–875.

Jian, J.-Y., Bisantz, A. M., and Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International journal of cognitive ergonomics*, 4(1):53–71.

Jing, L., Shan, W., and Zhang, Y. (2023). Risk preference, risk perception as predictors of risky driving behaviors: the moderating effects of gender, age, and driving experience. *Journal of Transportation Safety & Security*, 15(5):467–492.

Jokinen, J. P., Kujala, T., and Oulasvirta, A. (2021). Multitasking in driving as optimal adaptation under uncertainty. *Human Factors*, 63(8):1324–1341.

Kahneman, D. (1973). *Attention and effort*, volume 1063. Citeseer.

Karle, J. W., Watter, S., and Shedden, J. M. (2010). Task switching in video game players: Benefits of selective attention but not resistance to proactive interference. *Acta psychologica*, 134(1):70–78.

Karpinsky, N. D., Chancey, E. T., Palmer, D. B., and Yamani, Y. (2018). Automation trust and attention allocation in multitasking workspace. *Applied ergonomics*, 70:194–201.

Karthaus, M., Wascher, E., Falkenstein, M., and Getzmann, S. (2020). The ability of young, middle-aged and older drivers to inhibit visual and auditory distraction in a driving simulator task. *Transportation research part F: traffic psychology and behaviour*, 68:272–284.

Kashevnik, A., Shchedrin, R., Kaiser, C., and Stocker, A. (2021). Driver distraction detection methods: A literature review and framework. *IEEE Access*, 9:60063–60076.

Keller, A. S., Payne, L., and Sekuler, R. (2017). Characterizing the roles of alpha and theta oscillations in multisensory attention. *Neuropsychologia*, 99:48–63.

Kelly, S. P., Lalor, E. C., Reilly, R. B., and Foxe, J. J. (2006). Increases in alpha oscillatory power reflect an active retinotopic mechanism for distracter suppression during sustained visuospatial attention. *Journal of neurophysiology*, 95(6):3844–3851.

Khosravi, S., Li, H., Khan, A. R., Zoha, A., and Ghannam, R. (2024). Multi-modal-sensing system for detection and tracking of mind wandering. *Multimodal Intelligent Sensing* *in Modern Applications*, pages 181–200.

Kida, T., Kaneda, T., and Nishihira, Y. (2012). Dual-task repetition alters event-related brain potentials and task performance. *Clinical neurophysiology*, 123(6):1123–1130.

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., and Koch, I. (2010). Control and interference in task switching—a review. *Psychological bulletin*, 136(5):849.

Kievit, R. A., Davis, S. W., Mitchell, D. J., Taylor, J. R., Duncan, J., and Henson, R. N. (2014). Distinct aspects of frontal lobe structure mediate age-related differences in fluid intelligence and multitasking. *Nature communications*, 5(1):5658.

Kim, D.-W. and Im, C.-H. (2018). Eeg spectral analysis. *Computational EEG analysis: Methods and applications*, pages 35–53.

Kim, J. E., Nembhard, D. A., and Kim, J. H. (2016). The effects of group size and task complexity on deadline reactivity. *International Journal of Industrial Ergonomics*, 56:106–114.

Kimura, M., Kimura, K., and Takeda, Y. (2022). Assessment of driver's attentional resource allocation to visual, cognitive, and action processing by brain and eye signals. *Transportation research part F: traffic psychology and behaviour*, 86:161–177.

Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., and Dingus, T. A. (2014). Distracted driving and risk of road crashes among novice and experienced drivers. *New England journal of medicine*, 370(1):54–59.

Klimesch, W. (1999). Eeg alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews*, 29(2-3):169–195.

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, 16(12):606–617.

Klimesch, W. (2018). The frequency architecture of brain and brain body oscillations: an analysis. *European Journal of Neuroscience*, 48(7):2431–2453.

Koch, I., Poljac, E., Müller, H., and Kiesel, A. (2018). Cognitive structure, flexibility, and plasticity in human multitasking—an integrative review of dual-task and task-switching research. *Psychological bulletin*, 144(6):557.

Kong, Y., Posada-Quintero, H. F., Gever, D., Bonacci, L., Chon, K. H., and Bolkhovsky,

J. (2022). Multi-attribute task battery configuration to effectively assess pilot performance deterioration during prolonged wakefulness. *Informatics in Medicine Unlocked*, 28:100822.

Kountouriotis, G. K. and Merat, N. (2016). Leading to distraction: Driver distraction, lead car, and road environment. *Accident Analysis & Prevention*, 89:22–30.

Kredel, R., Vater, C., Klostermann, A., and Hossner, E.-J. (2017). Eye-tracking technology and the dynamics of natural gaze behavior in sports: A systematic review of 40 years of research. *Frontiers in psychology*, 8:1845.

Kulke, L., Atkinson, J., and Braddick, O. (2015). Automatic detection of attention shifts in infancy: Eye tracking in the fixation shift paradigm. *PLoS ONE*, 10(12):e0142505.

Lamb, S. and Kwok, K. C. (2016). A longitudinal investigation of work environment stressors on the performance and wellbeing of office workers. *Applied Ergonomics*, 52:104–111.

Lan, T., Hu, H., Jiang, C., Yang, G., and Zhao, Z. (2020). A comparative study of decision tree, random forest, and convolutional neural network for spread-f identification. *Advances in Space Research*, 65(8):2052–2061.

Lee, B. C. and Duffy, V. G. (2015). The effects of task interruption on human performance: A study of the systematic classification of human behavior and interruption frequency. *Human Factors and Ergonomics in Manufacturing Service Industries*, 25(2):137–152.

Lee, J., Wickens, C., Liu, Y., and Boyle, L. (2017). Designing for people. *ShangHai*, page 173.

Lee, J. D. and See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human factors*, 46(1):50–80.

Lee, J. D., Young, K. L., and Regan, M. A. (2008). Defining driver distraction. *Driver distraction: Theory, effects, and mitigation*, 13(4):31–40.

Lehrer, P., Karavidas, M., Lu, S.-E., Vaschillo, E., Vaschillo, B., and Cheng, A. (2010). Cardiac data increase association between self-report and both expert ratings of task load and task performance in flight simulator tasks: An exploratory study. *International Journal of Psychophysiology*, 76(2):80–87.

Lemonnier, S., Désiré, L., Brémond, R., and Baccino, T. (2020). Drivers' visual attention: A field study at intersections. *Transportation research part F: traffic psychology and* *behaviour*, 69:206–221.

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., and Liu, H. (2017). Feature selection: A data perspective. *ACM computing surveys (CSUR)*, 50(6):1–45.

Li, J., Huang, S., and Kim, J.-E. (2025). Predicting multitasking performance: An eeg-and eye movement-based dynamic bayesian network. *International Journal of Human–Computer Interaction*, pages 1–11.

Li, J. and Kim, J.-E. (2021). The effect of task complexity on time estimation in the virtual reality environment: an eeg study. *Applied Sciences*, 11(20):9779.

Li, W., Li, R., Xie, X., and Chang, Y. (2022). Evaluating mental workload during multitasking in simulated flight. *Brain and Behavior*, 12(4):e2489.

Li, X. and Ji, Q. (2004). Active affective state detection and user assistance with dynamic bayesian networks. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 35(1):93–105.

Liao, W., Zhang, W., Zhu, Z., and Ji, Q. (2005). A real-time human stress monitoring system using dynamic bayesian network. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, page 70. IEEE.

Lim, W., Sourina, O., and Wang, L. P. (2018). Stew: Simultaneous task eeg workload data set. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(11):2106–2114.

Lin, C.-T., Chen, S.-A., Chiu, T.-T., Lin, H.-Z., and Ko, L.-W. (2011). Spatial and temporal eeg dynamics of dual-task driving performance. *Journal of neuroengineering and rehabilitation*, 8(1):1–13.

Lin, L. (2013). Multiple dimensions of multitasking phenomenon. *International Journal of Technology and Human Interaction (IJTHI)*, 9(1):37–49.

Liu, P. and Li, Z. (2012). Task complexity: A review and conceptualization framework. *International Journal of Industrial Ergonomics*, 42(6):553–568.

Liu, S. and Nam, C. S. (2018). Quantitative modeling of user performance in multitasking environments. *Computers in Human Behavior*, 84:130–140.

Liu, S., Wadeson, A., Kim, N. Y., and Nam, C. S. (2016). Effects of working memory capacity, task switching, and task difficulty on multitasking performance. In *Proceed-*

ings of the Human Factors and Ergonomics Society annual meeting, volume 60, pages 502–506. SAGE Publications Sage CA: Los Angeles, CA.

Liu, Y. and Gao, Q. (2024). Effects of secondary task eccentricity and visual salience on attention allocation in multitasking across screens. *International Journal of Human-Computer Studies*, 192:103363.

Lui, K. F. and Wong, A. C.-N. (2020). Multiple processing limitations underlie multitasking costs. *Psychological research*, 84(7):1946–1964.

Lyell, D. and Coiera, E. (2017). Automation bias and verification complexity: a systematic review. *Journal of the American Medical Informatics Association*, 24(2):423–431.

Machado-León, J. L., de Oña, J., de Oña, R., Eboli, L., and Mazzulla, G. (2016). Socio-economic and driving experience factors affecting drivers' perceptions of traffic crash risk. *Transportation research part F: traffic psychology and behaviour*, 37:41–51.

MacQueen, J. (1967). Classification and analysis of multivariate observations. In *5th Berkeley Symp. Math. Statist. Probability*, pages 281–297.

Mark, G., Gonzalez, V. M., and Harris, J. (2005). No task left behind? examining the nature of fragmented work. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 321–330.

Mark, J., Curtin, A., Kraft, A., Sargent, A., Perez, A., Friedman, L., Barkan, A., Sands, T., Casebeer, W., Ziegler, M., et al. (2018). Multimodal cognitive workload assessment using eeg, fnirs, ecg, eog, ppg, and eye-tracking. *Front. Hum. Neurosci*, 12(2018):10–3389.

Martens, M. H. and Fox, M. (2007). Does road familiarity change eye fixations? a comparison between watching a video and real driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 10(1):33–47.

Martin, J., Mashburn, C. A., and Engle, R. W. (2020). Improving the validity of the armed service vocational aptitude battery with measures of attention control. *Journal of Applied Research in Memory and Cognition*, 9(3):323–335.

McAvinue, L. P., Habekost, T., Johnson, K. A., Kyllingsbæk, S., Vangkilde, S., Bundesen, C., and Robertson, I. H. (2012). Sustained attention, attentional selectivity, and attentional capacity across the lifespan. *Attention, Perception, & Psychophysics*, 74:1570–1582.

McCartt, A. T., Hellinga, L. A., and Bratiman, K. A. (2006). Cell phones and driving: review of research. *Traffic injury prevention*, 7(2):89–106.

McEvoy, L. K., Pellouchoud, E., Smith, M. E., and Gevins, A. (2001). Neurophysiological signals of working memory in normal aging. *Cognitive Brain Research*, 11(3):363–376.

Meng, Q., An, Y., and Yang, D. (2021). Effects of acoustic environment on design work performance based on multitask visual cognitive performance in office space. *Building and Environment*, 205:108296.

Metzger, U. and Parasuraman, R. (2017). Automation in future air traffic management: Effects of decision aid reliability on controller performance and mental workload. In *Decision Making in Aviation*, pages 345–360. Routledge.

Meyer, D. E. and Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part i. basic mechanisms. *Psychological review*, 104(1):3.

Mihajlovic, V. and Petkovic, M. (2001). Dynamic bayesian networks: A state of the art. Technical report, University of Twente Document Repository.

Minear, M., Brasher, F., McCurdy, M., Lewis, J., and Younggren, A. (2013). Working memory, fluid intelligence, and impulsiveness in heavy media multitaskers. *Psychonomic bulletin & review*, 20:1274–1281.

Minear, M. and Shah, P. (2008). Training and transfer effects in task switching. *Memory & cognition*, 36:1470–1483.

Moacdieh, N. M., Devlin, S. P., Jundi, H., and Riggs, S. L. (2020). Effects of workload and workload transitions on attention allocation in a dual-task environment: Evidence from eye tracking metrics. *Journal of Cognitive Engineering and Decision Making*, 14(2):132–151.

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

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2021). *Introduction to Linear Regression Analysis*. John Wiley Sons.

Moreira, P. E. D., Dieguez, G. T. d. O., Bredt, S. d. G. T., and Praça, G. M. (2021). The acute and chronic effects of dual-task on the motor and cognitive performances in athletes: a systematic review. *International Journal of Environmental Research and Public Health*, 18(4):1732.

Morgan, B., D'Mello, S., Abbott, R., Radvansky, G., Haass, M., and Tamplin, A. (2013). Individual differences in multitasking ability and adaptability. *Human factors*, 55(4):776–788.

Mundell, C., Vielma, J. P., and Zaman, T. (2016). Predicting performance under stressful conditions using galvanic skin response. *arXiv preprint arXiv:1606.01836*.

Murphy, K. P. et al. (2002). Dynamic bayesian networks. *Probabilistic Graphical Models, M. Jordan*, 7:431.

Murray, M. M., De Santis, L., Thut, G., and Wylie, G. R. (2009). The costs of crossing paths and switching tasks between audition and vision. *Brain and cognition*, 69(1):47–55.

Murugappan, M., Murugappan, S., and Gerard, C. (2014). Wireless eeg signals based neuromarketing system using fast fourier transform (fft). In *2014 IEEE 10th International Colloquium on Signal Processing and its Applications*, pages 25–30. IEEE.

Mäntylä, T. (2013). Gender differences in multitasking reflect spatial ability. *Psychological Science*, 24(4):514–520.

Nagel, B. J., Herting, M. M., Maxwell, E. C., Bruno, R., and Fair, D. (2013). Hemispheric lateralization of verbal and spatial working memory during adolescence. *Brain and cognition*, 82(1):58–68.

Nahm, F. S. (2016). Nonparametric statistical tests for the continuous data: the basic concept and the practical use. *Korean journal of anesthesiology*, 69(1):8–14.

Navon, D. and Gopher, D. (1979). On the economy of the human-processing system. *Psychological review*, 86(3):214.

Neapolitan, R. E. and Jiang, X. (2010). *Probabilistic methods for financial and marketing informatics*. Elsevier.

Neth, H., Khemlani, S. S., and Gray, W. D. (2008). Feedback design for the control of a dynamic multitasking system: Dissociating outcome feedback from control feedback. *Human factors*, 50(4):643–651.

Neyens, D. M. and Boyle, L. N. (2008). The influence of driver distraction on the severity of injuries sustained by teenage drivers and their passengers. *Accident Analysis &*

Prevention, 40(1):254–259.

Nidamanuri, J., Nibhanupudi, C., Assfalg, R., and Venkataraman, H. (2021). A progressive review: Emerging technologies for adas driven solutions. *IEEE Transactions on Intelligent Vehicles*, 7(2):326–341.

Nijboer, M., Taatgen, N. A., Brands, A., Borst, J. P., and van Rijn, H. (2013). Decision making in concurrent multitasking: do people adapt to task interference? *PloS one*, 8(11):e79583.

Noble, A. M., Miles, M., Perez, M. A., Guo, F., and Klauer, S. G. (2021). Evaluating driver eye glance behavior and secondary task engagement while using driving automation systems. *Accident Analysis & Prevention*, 151:105959.

Nodelman, U., Shelton, C. R., and Koller, D. (2012). Learning continuous time bayesian networks. arXiv preprint arXiv:1212.2498.

O'Leary, D. S., Andreasen, N., Hurtig, R., Torres, I., Flashman, L., Kesler, M., Arndt, S., Cizadlo, T., Ponto, L., Watkins, G., et al. (1997). Auditory and visual attention assessed with pet. *Human brain mapping*, 5(6):422–436.

Onnasch, L. (2015). Crossing the boundaries of automation—function allocation and reliability. *International Journal of Human-Computer Studies*, 76:12–21.

Onnasch, L., Wickens, C. D., Li, H., and Manzey, D. (2014). Human performance consequences of stages and levels of automation: An integrated meta-analysis. *Human factors*, 56(3):476–488.

Osman, O. A., Hajij, M., Karbalaieali, S., and Ishak, S. (2019). A hierarchical machine learning classification approach for secondary task identification from observed driving behavior data. *Accid. Anal. Prev.*, 123:274–281.

Oviedo-Trespalacios, O., Haque, M. M., King, M., and Washington, S. (2016). Understanding the impacts of mobile phone distraction on driving performance: A systematic review. *Transportation research part C: emerging technologies*, 72:360–380.

Ozdemir, R. A., Contreras-Vidal, J. L., Lee, B. C., and Paloski, W. H. (2016). Cortical activity modulations underlying age-related performance differences during posture–cognition dual tasking. *Experimental Brain Research*, 234(11):3321–3334.

O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors.

Quality & quantity, 41:673–690.

Pantangi, S. S., Ahmed, S. S., Fountas, G., Majka, K., and Anastasopoulos, P. C. (2021). Do high visibility crosswalks improve pedestrian safety? a correlated grouped random parameters approach using naturalistic driving study data. *Analytic methods in accident research*, 30:100155.

Papantoniou, P., Papadimitriou, E., and Yannis, G. (2017). Review of driving performance parameters critical for distracted driving research. *Transportation research procedia*, 25:1796–1805.

Pape, A. M., Wiegmann, D. A., and Shappell, S. A. (2001). Air traffic control (atc) related accidents and incidents: A human factors analysis.

Papulová, Z., Gažová, A., and Šufliarskỳ, L. (2022). Implementation of automation technologies of industry 4.0 in automotive manufacturing companies. *Procedia Computer Science*, 200:1488–1497.

Parasuraman, R. (2003). Neuroergonomics: research and practice. *Theoretical Issues in Ergonomics Science*, 4(1-2):5–20.

Parasuraman, R. and Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human factors*, 52(3):381–410.

Parasuraman, R., Sheridan, T. B., and Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, 30(3):286–297.

Parasuraman, R. and Wickens, C. D. (2017). Humans: Still vital after all these years of automation. In *Decision Making in Aviation*, pages 251–260. Routledge.

Paridon, H. M. and Kaufmann, M. (2010). Multitasking in work-related situations and its relevance for occupational health and safety: Effects on performance, subjective strain and physiological parameters. *Europe's Journal of Psychology*, 6(4):110–124.

Pashler, H. (1984). Processing stages in overlapping tasks: evidence for a central bottleneck. *Journal of Experimental Psychology: Human perception and performance*, 10(3):358.

Pashler, H. (1993). Doing two things at the same time. *American Scientist*, 81(1):48–55.

Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychological bulletin*, 116(2):220.

Peißl, S., Wickens, C. D., and Baruah, R. (2018). Eye-tracking measures in aviation: A selective literature review. *The International Journal of Aerospace Psychology*, 28(3-4):98–112.

Peng, A., Kirkham, N. Z., and Mareschal, D. (2018). Information processes of task-switching and modality-shifting across development. *PLoS One*, 13(6):e0198870.

Pew, R. W. and Mavor, A. S. (1998). Modeling human and organizational behavior-application to military simulations.

Pike, A. M., Shirinzad, M., Ananda, R. S., and Pawar, A. R. (2022). Gaze behaviors of drivers approaching crosswalks with different sign and crosswalk treatments. *Transportation Research Record*, page 03611981221132855.

Plummer, P. and Eskes, G. (2015). Measuring treatment effects on dual-task performance: a framework for research and clinical practice. *Frontiers in human neuroscience*, 9:225.

Pollard, M. A. and Courage, M. L. (2017). Working memory capacity predicts effective multitasking. *Computers in human behavior*, 76:450–462.

Pontiggia, A., Gomez-Merino, D., Quiquempoix, M., Beauchamps, V., Boffet, A., Fabries, P., Chennaoui, M., and Sauvet, F. (2024). Matb for assessing different mental workload levels. *Frontiers in Physiology*, 15:1408242.

Puma, S., Matton, N., Paubel, P.-V., Raufaste, É., and El-Yagoubi, R. (2018). Using theta and alpha band power to assess cognitive workload in multitasking environments. *International Journal of Psychophysiology*, 123:111–120.

Qin, L., Li, Z. R., Chen, Z., Bill, M. A., and Noyce, D. A. (2019). Understanding driver distractions in fatal crashes: An exploratory empirical analysis. *Journal of Safety Research*, 69:23–31.

R Core Team (2020). R: A language and environment for statistical computing. `https://www.R-project.org/`.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reason, J. (1990). *Human error*. Cambridge University Press.

Regan, M. A., Hallett, C., and Gordon, C. P. (2011). Driver distraction and driver inattention: Definition, relationship and taxonomy. *Accident Analysis & Prevention*, 43(5):1771–1781.

Regan, M. A., Lee, J. D., and Young, K. (2008). *Driver distraction: Theory, effects, and mitigation*. CRC press.

Reichenberg, R. (2018). Dynamic bayesian networks in educational measurement: Reviewing and advancing the state of the field. *Applied Measurement in Education*, 31(4):335–350.

Reimers, S. and Maylor, E. A. (2005). Task switching across the life span: effects of age on general and specific switch costs. *Developmental psychology*, 41(4):661.

Ren, D., Zhou, H., and Fu, X. (2009). A deeper look at gender difference in multitasking: Gender-specific mechanism of cognitive control. In *2009 Fifth International Conference on Natural Computation*, volume 5, pages 13–17. IEEE.

Rhodes, N. and Pivik, K. (2011). Age and gender differences in risky driving: The roles of positive affect and risk perception. *Accident Analysis & Prevention*, 43(3):923–931.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied linguistics*, 22(1):27–57.

Rosenfeld, J. P., Ward, A., Meijer, E. H., and Yukhnenko, D. (2017). Bootstrapping the p300 in diagnostic psychophysiology: How many iterations are needed? *Psychophysiology*, 54(3):366–373.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.

Różańska, A. and Gruszka, A. (2020). Current research trends in multitasking: a bibliometric mapping approach. *Journal of Cognitive Psychology*, 32(3):278–286.

Rubinstein, J. S., Meyer, D. E., and Evans, J. E. (2001). Executive control of cognitive processes in task switching. *Journal of experimental psychology: human perception and performance*, 27(4):763.

Sajid Hasan, A., Jalayer, M., Heitmann, E., and Weiss, J. (2022). Distracted driving crashes: a review on data collection, analysis, and crash prevention methods. *Transportation research record*, 2676(8):423–434.

Salah, J., Abdelrahman, Y., Abdrabou, Y., Kassem, K., and Abdennadher, S. (2018). Exploring the usage of commercial bio-sensors for multitasking detection. In *Proceedings* *of the 17th International Conference on Mobile and Ubiquitous Multimedia*, pages 265–277.

Salvucci, D. D. and Taatgen, N. A. (2008). Threaded cognition: an integrated theory of concurrent multitasking. *Psychological review*, 115(1):101.

Sanbonmatsu, D. M., Strayer, D. L., Medeiros-Ward, N., and Watson, J. M. (2013). Who multi-tasks and why? multi-tasking ability, perceived multi-tasking ability, impulsivity, and sensation seeking. *PloS one*, 8(1):e54402.

Santiago-Espada, Y., Myer, R. R., Latorella, K. A., and Comstock Jr, J. R. (2011). The multi-attribute task battery ii (matb-ii) software for human performance and workload research: A user's guide. Technical Report L-20031, NASA.

Sarwar, M. T., Fountas, G., Bentley, C., Anastasopoulos, P. C., Blatt, A., Pierowicz, J., Majka, K., and Limoges, R. (2017). Preliminary investigation of the effectiveness of high-visibility crosswalks on pedestrian safety using crash surrogates. *Transportation Research Record*, 2659(1):182–191.

Sato, T., Inman, J., Politowicz, M. S., Chancey, E. T., and Yamani, Y. (2023a). A meta-analytic approach to investigating the relationship between human-automation trust and attention allocation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 67, pages 959–964. SAGE Publications Sage CA: Los Angeles, CA.

Sato, T., Islam, S., Still, J. D., Scerbo, M. W., and Yamani, Y. (2023b). Task priority reduces an adverse effect of task load on automation trust in a dynamic multitasking environment. *Cognition, Technology & Work*, 25(1):1–13.

Savage, S. W., Potter, D. D., and Tatler, B. W. (2013). Does preoccupation impair hazard perception? a simultaneous eeg and eye tracking study. *Transportation research part F: traffic psychology and behaviour*, 17:52–62.

Sazgar, M. and Young, M. G. (2019). Overview of eeg, electrode placement, and montages. In *Absolute Epilepsy and EEG Rotation Review*, pages 117–125. Springer, Cham.

Schläpfer, J. and Wellens, H. J. (2017). Computer-interpreted electrocardiograms: benefits and limitations. *Journal of the American College of Cardiology*, 70(9):1183–1192.

Schneider, R. J. and Sanders, R. L. (2015). Pedestrian safety practitioners' perspectives of driver yielding behavior across north america. *Transportation Research Record*, 2519(1):39–50.

Schroeder, P., Wilbur, M., Peña, R., et al. (2018). National survey on distracted driving attitudes and behaviors-2015. Technical report, United States. National Highway Traffic Safety Administration.

Sciaraffa, N., Di Flumeri, G., Germano, D., Giorgi, A., Di Florio, A., Borghini, G., Vozzi, A., Ronca, V., Varga, R., van Gasteren, M., et al. (2022). Validation of a light eeg-based measure for real-time stress monitoring during realistic driving. *Brain sciences*, 12(3):304.

Scutari, M. (2009). Learning bayesian networks with the bnlearn r package. arXiv preprint arXiv:0908.3817.

Sebok, A., Wickens, C., Sarter, N., Quesada, S., Socash, C., and Anthony, B. (2012). The automation design advisor tool (adat): Development and validation of a model-based tool to support flight deck automation design for nextgen operations. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 22(5):378–394.

Sevil, M., Rashid, M., Askari, M. R., Maloney, Z., Hajizadeh, I., and Cinar, A. (2020). Detection and characterization of physical activity and psychological stress from wristband data. *Signals*, 1(2):188–208.

Sheridan, T. B., Verplank, W. L., and Brooks, T. (1978). Human/computer control of undersea teleoperators. In *NASA. Ames Res. Center The 14th Ann. Conf. on Manual Control*.

Shiferaw, B. A., Downey, L. A., Westlake, J., Stevens, B., Rajaratnam, S. M., Berlowitz, D. J., Swann, P., and Howard, M. E. (2018). Stationary gaze entropy predicts lane departure events in sleep-deprived drivers. *Scientific reports*, 8(1):2220.

Shinar, D., Schechtman, E., and Compton, R. (2001). Self-reports of safe driving behaviors in relationship to sex, age, education and income in the us adult driving population. *Accident Analysis & Prevention*, 33(1):111–116.

Sinaga, K. P. and Yang, M.-S. (2020). Unsupervised k-means clustering algorithm. *IEEE access*, 8:80716–80727.

Song, Y. Y. and Ying, L. U. (2015). Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry*, 27(2):130.

Spink, A., Cole, C., and Waller, M. (2008). Multitasking behavior. *Annual review of information science and technology*, 42(1):93–118.

Stavrinos, D., McManus, B., and Beck, H. (2020). Demographic, driving experience, and psychosocial predictors of adolescent distracted driving beliefs. *Accident Analysis & Prevention*, 144:105678.

Stavrinos, D., Pope, C. N., Shen, J., and Schwebel, D. C. (2018). Distracted walking, bicycling, and driving: Systematic review and meta-analysis of mobile technology and youth crash risk. *Child development*, 89(1):118–128.

Stephan, D. N. and Koch, I. (2011). The role of input–output modality compatibility in task switching. *Psychological Research*, 75(6):491–498.

Stock, A.-K., Gohil, K., Huster, R. J., and Beste, C. (2017). On the effects of multimodal information integration in multitasking. *Scientific Reports*, 7(1):4927.

Stoet, G., O'Connor, D. B., Conner, M., and Laws, K. R. (2013). Are women better than men at multi-tasking? *BMC Psychology*, 1(1):1–10.

Stoica, P. and Selen, Y. (2004). Model-order selection: a review of information criterion rules. *IEEE signal processing magazine*, 21(4):36–47.

Strybel, T. Z., Vu, K.-P. L., Chiappe, D. L., Morgan, C. A., Morales, G., and Battiste, V. (2016). Effects of nextgen concepts of operation for separation assurance and interval management on air traffic controller situation awareness, workload, and performance. *The International Journal of Aviation Psychology*, 26(1-2):1–14.

Svetina, M. (2016). The reaction times of drivers aged 20 to 80 during a divided attention driving. *Traffic injury prevention*, 17(8):810–814.

Szandała, T. (2021). Review and comparison of commonly used activation functions for deep neural networks. In *Bio-inspired Neurocomputing*, pages 203–224. Springer, Singapore.

Szumowska, E. and Kossowska, M. (2016). Need for closure and multitasking performance: The role of shifting ability. *Personality and Individual Differences*, 96:12–17.

Tang, J. (2017). Analysis and improvement of traffic alert and collision avoidance system. *IEEE Access*, 5:21419–21429.

Tao, D., Tan, H., Wang, H., Zhang, X., Qu, X., and Zhang, T. (2019). A systematic review of physiological measures of mental workload. *International Journal of Environmental Research and Public Health*, 16(15):2716.

Telford, C. W. (1931). The refractory phase of voluntary and associative responses. *Journal of Experimental Psychology*, 14(1):1.

Teodorovicz, T., Kun, A. L., Sadun, R., and Shaer, O. (2022). Multitasking while driving: A time use study of commuting knowledge workers to assess current and future uses. *International Journal of Human-Computer Studies*, 162:102789.

Todorov, I., Del Missier, F., and Mäntylä, T. (2014). Age-related differences in multiple task monitoring. *PLoS One*, 9(9):e107619.

Tsai, Y.-F., Viirre, E., Strychacz, C., Chase, B., and Jung, T.-P. (2007). Task performance and eye activity: predicting behavior relating to cognitive workload. *Aviation, space, and environmental medicine*, 78(5):B176–B185.

Tsang, S. N. H., Chan, A. H. S., Pan, X., and Man, S. S. (2021). Auditory versus visual spatial stimulus-response mappings in tracking and discrete dual task performance: Implications for human-machine interface design. *Ergonomics*, 64(4):485–501.

Valéry, B., Matton, N., Scannella, S., and Dehais, F. (2019). Global difficulty modulates the prioritization strategy in multitasking situations. *Applied Ergonomics*, 80:1–8.

Van Rossum, G. and Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA.

Vansteenkiste, P., Cardon, G., Philippaerts, R., and Lenoir, M. (2015). Measuring dwell time percentage from head-mounted eye-tracking data–comparison of a frame-by-frame and a fixation-by-fixation analysis. *Ergonomics*, 58(5):712–721.

Varhelyi, A. (1998). Drivers' speed behaviour at a zebra crossing: a case study. *Accident Analysis & Prevention*, 30(6):731–743.

Verhaeghen, P., Steitz, D. W., Sliwinski, M. J., and Cerella, J. (2003). Aging and dual-task performance: a meta-analysis. *Psychology and aging*, 18(3):443.

Vidulich, M. A. and Tsang, P. S. (2012). Mental workload and situation awareness. *Handbook of human factors and ergonomics*, pages 243–273.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). Scipy 1.0: fundamental algorithms for scientific computing in python.

Nature Methods, 17(3):261–272.

Volf, P., Jakubuv, J., Koranda, L., Sislák, D., Pěchoucek, M., Mereu, S., Hilburn, B., and Nguyen, D. N. (2014). Validation of an air-traffic controller behavioral model for fast time simulation. In *2014 Integrated Communications, Navigation and Surveillance Conference (ICNS) Conference Proceedings*, pages T1–1. IEEE.

Vu, K.-P. L., Strybel, T. Z., Battiste, V., Lachter, J., Dao, A.-Q. V., Brandt, S., Ligda, S., and Johnson, W. (2012). Pilot performance in trajectory-based operations under concepts of operation that vary separation responsibility across pilots, air traffic controllers, and automation. *International Journal of Human-Computer Interaction*, 28(2):107–118.

Wahn, B., Murali, S., Sinnett, S., and Koenig, P. (2017). Auditory stimulus detection partially depends on visuospatial attentional resources. *i-Perception*, 8(1):2041669516688026.

Wallach, D. and Goffinet, B. (1989). Mean squared error of prediction as a criterion for evaluating and comparing system models. *Ecological Modelling*, 44(3-4):299–306.

Walter, S. R., Li, L., Dunsmuir, W. T., and Westbrook, J. I. (2013). Managing competing demands through task-switching and multitasking: a multi-setting observational study of 200 clinicians over 1000 hours. *BMJ quality & safety*.

Wang, Y., Hu, R., Lin, S., Schultz, M., and Delahaye, D. (2021). The impact of automation on air traffic controller's behaviors. *Aerospace*, 8(9):260.

Wang, Z., David, P., Srivastava, J., Powers, S., Brady, C., D'Angelo, J., and Moreland, J. (2012). Behavioral performance and visual attention in communication multitasking: A comparison between instant messaging and online voice chat. *Computers in Human Behavior*, 28(3):968–975.

Waszak, F., Hommel, B., and Allport, A. (2003). Task-switching and long-term priming: Role of episodic stimulus–task bindings in task-shift costs. *Cognitive psychology*, 46(4):361–413.

Wechsler, K., Drescher, U., Janouch, C., Haeger, M., Voelcker-Rehage, C., and Bock, O. (2018). Multitasking during simulated car driving: a comparison of young and older persons. *Frontiers in psychology*, 9:910.

Weigl, M., Müller, A., Sevdalis, N., and Angerer, P. (2013). Relationships of multitasking, physicians' strain, and performance. *Journal of patient safety*, 9(1):18–23.

Welch, P. (1967). The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73.

Welford, A. (1959). Evidence of a single-channel decision mechanism limiting performance in a serial reaction task. *Quarterly Journal of Experimental Psychology*, 11(4):193–210.

Werneke, J. and Vollrath, M. (2012). What does the driver look at? the influence of intersection characteristics on attention allocation and driving behavior. *Accident Analysis & Prevention*, 45:610–619.

Westbrook, J. I., Raban, M. Z., Walter, S. R., and Douglas, H. (2018). Task errors by emergency physicians are associated with interruptions, multitasking, fatigue and working memory capacity: a prospective, direct observation study. *BMJ quality & safety*, 27(8):655–663.

Wetherell, M. A. and Carter, K. (2014). The multitasking framework: The effects of increasing workload on acute psychobiological stress reactivity. *Stress and Health*, 30(2):103–109.

Wetherell, M. A., Craw, O., Smith, K., and Smith, M. A. (2017). Psychobiological responses to critically evaluated multitasking. *Neurobiology of stress*, 7:68–73.

Wickens, C. (2021). Attention: Theory, principles, models and applications. *International Journal of Human–Computer Interaction*, 37(5):403–417.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, 3(2):159–177.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human factors*, 50(3):449–455.

Wickens, C. D. (2020). Processing resources and attention. In *Multiple task performance*, pages 3–34. CRC Press.

Wickens, C. D., Gutzwiller, R. S., and 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., Helton, W. S., Hollands, J. G., and Banbury, S. (2021). *Engineering psychology and human performance*. Routledge.

Wiczorek, R. and Manzey, D. (2014). Supporting attention allocation in multitask environments: Effects of likelihood alarm systems on trust, behavior, and performance. *Human factors*, 56(7):1209–1221.

Williamson, T. and Spencer, N. A. (1989). Development and operation of the traffic alert and collision avoidance system (tcas). *Proceedings of the IEEE*, 77(11):1735–1744.

Wilson, G. F. and Eggemeier, F. T. (2020). *Psychophysiological assessment of workload in multi-task environments*. Taylor Francis.

Wojton, H. M., Porter, D., T. Lane, S., Bieber, C., and Madhavan, P. (2020). Initial validation of the trust of automated systems test (toast). *The Journal of social psychology*, 160(6):735–750.

Wong, T. T. (2015). Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition*, 48(9):2839–2846.

Wood, R. E. (1986). Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes*, 37(1):60–82.

Yamani, Y. and Horrey, W. J. (2018). A theoretical model of human-automation interaction grounded in resource allocation policy during automated driving. *International Journal of Human Factors and Ergonomics*, 5(3):225–239.

Yang, G., Lin, Y., and Bhattacharya, P. (2010). A driver fatigue recognition model based on information fusion and dynamic bayesian network. *Information Sciences*, 180(10):1942–1954.

Ye, M., Osman, O. A., and Ishak, S. (2017). Accounting for driver distraction and socioeconomic characteristics in a crash risk index: naturalistic driving study. *Transportation Research Record*, 2659(1):204–211.

Yoshizawa, A. and Iwasaki, H. (2015). Influence of nonvisual secondary tasks on driver's pedestrian detection. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, 9(4):21–32.

Young, K., Regan, M., and Hammer, M. (2007). Driver distraction: A review of the literature. *Distracted driving*, 2007:379–405.

Young, K. L. and Lenné, M. G. (2010). Driver engagement in distracting activities and the strategies used to minimise risk. *Safety science*, 48(3):326–332.

Young, K. L., Osborne, R., Grzebieta, R., Williamson, A., Haworth, N., Senserrick, T., Stephan, K. L., and Charlton, J. L. (2020). Using naturalistic driving data to examine how drivers share attention when engaging in secondary tasks. *Safety science*, 129:104841.

Young, K. L., Regan, M. A., and Lee, J. D. (2009). Measuring the effects of driver distraction: Direct driving performance methods and measures. *Driver distraction: Theory, effects and mitigation*, pages 85–106.

Zegeer, C. V., Stewart, J. R., Huang, H. H., Lagerwey, P. A., Feaganes, J. R., Campbell, B., et al. (2005). Safety effects of marked versus unmarked crosswalks at uncontrolled locations final report and recommended guidelines. Technical report, United States. Federal Highway Administration. Office of Safety Research and . . . .

Zeng, Z. and Ji, Q. (2010). Knowledge based activity recognition with dynamic bayesian network. In *European Conference on Computer Vision*, pages 532–546, Berlin, Heidelberg. Springer.

Zhang, Y., Kaber, D. B., Rogers, M., Liang, Y., and Gangakhedkar, S. (2014). The effects of visual and cognitive distractions on operational and tactical driving behaviors. *Human factors*, 56(3):592–604.

Zou, T., Guo, H., Khaloei, M., MacKenzie, D., and Boyle, L. N. (2023). Examining the relationships between multimodal environments and multitasking driving behaviors. *Transp. Res. Rec.*, 2677(2):944–957.