QUANTIFYING COGNITIVE EFFICIENCY OF DISPLAY IN HUMAN-MACHINE SYSTEMS

A Dissertation

by

SHIYAN YANG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Thomas K. Ferris
Committee Members,	Steve Smith
	Ricardo Gutierrez-Osuna
	Yu Ding
Head of Department,	César O. Malavé

August 2016

Major Subject: Industrial Engineering

ABSTRACT

As a side effect of fast growing informational technology, information overload becomes prevalent in the operation of many human-machine systems. Overwhelming information can degrade operational performance because it imposes large mental workload on human operators. One way to address this issue is to improve the *cognitive efficiency* of display. A cognitively efficient display should be more informative while demanding less mental resources so that an operator can process larger displayed information using their limited working memory and achieve better performance. In order to quantitatively evaluate this display property, a Cognitive Efficiency (CE) metric is formulated as the ratio of the measures of two dimensions: *display informativeness* and *required mental resources* (each dimension can be affected by display, human, and contextual factors).

The first segment of the dissertation discusses the available measurement techniques to construct the CE metric and initially validates the CE metric with basic discrete displays. The second segment demonstrates that displays showing higher cognitive efficiency improve multitask performance. This part also identifies the version of the CE metric that is the most predictive of multitask performance. The last segment of the dissertation applies the CE metric in driving scenarios to evaluate novel speedometer displays; however, it finds that the most efficient display may not better enhance concurrent tracking performance in driving. Although the findings of dissertation show

several limitations, they provide valuable insight into the complicated relationship among display, human cognition, and multitask performance in human-machine systems.

DEDICATION

To Fiona

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Ferris, and my committee members, Dr. Smith, Dr. Gutierrez-Osuna, and Dr. Ding, for their guidance and support throughout the course of this research.

Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience. Especially, I would like to thank the PhD students in the HF&CS lab, Trey Roady, Katie Tippey, Johneen Ardoin, Youngbo Suh, and Carolina Rodrigues Paras, for their help during my PhD training.

Finally, thanks to my mother and father for their encouragement and to my wife for her patience and love.

NOMENCLATURE

CE	Cognitive Efficiency
HRV	Heart Rate Variability
SCR	Skin Conductance Response
EEG	Electroencephalogram

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

I.1 Information Overload

The operational success of complex human-machine systems largely depends on the machines' capacity to convey task-relevant information to human operators. For example, a well-designed car should be able to effectively present drivers with various types of information, such as speed on the speedometer, planned route on the GPS, traffic next to and behind in the rear view mirrors and abnormal vehicle states by the warning lights on the vehicle panel. The display design is critical for human operators to assess the ongoing state of human-machine systems and satisfactorily perform multiple concurrent tasks.

Influenced by the fast-growing computational technologies, the displays in modern human-machine systems tend to present more and more data in sophisticated formats (e.g., the control panel of Tesla electric car). Since human only has limited working memory (Simon, 1999), the displayed information could be more than what our cognitive systems can handle, thus causing a challenging problem: *information overload*. Besides overwhelming information, ineffective display design, such as those show clutter (Moacdieh & Sarter, 2015), may also result in information overload because it's hard for human to distribute attention precisely to the target message among rich data set on poorly-designed displays (Woods, Patterson, & Roth, 2002). Given the gap between exponentially-growing display technologies and ‘stone-age’ human brain, information overload becomes a generic problem that resists many solutions.

In addition to display and human factors, contextual factors can also contribute to information overload. In human-machine systems, the contextual factors mainly refer to physical energy that interrupt human attention, such as ambient lights and sounds. In larger sociotechnical systems, the contextual factors also include managerial factors, such as task procedure (e.g., too many steps of completing an operational task) and organizational design (e.g., heavy collaborative work among multiple operators) (Eppler & Mengis, 2004). Considering all relevant factors, information overload should be viewed systematically as an emergent phenomenon generated by the interaction of human, display, and contextual factors.

As a prevalent problem, information overload exists in a wide range of scenarios, including surface transportation, aviation, healthcare operation, and business management, being called in multiple names, such as data overload (Woods, Patterson, & Roth, 2002), knowledge overload (Hunt & Newman, 1997), and cognitive overload (Vollmann, 1991). However, for the sake of content consistency the dissertation stayed at using the name information overload.

Performance degradation is a critical negative consequence of information overload. According to the research of cognitive engineering, overwhelming information imposes a large amount of mental workload beyond the cognitive redline (e.g., the point when mental workload surpasses available mental resources), thus leading to worse task performance (Wickens, Hollands, Banbury, & Parasuraman., 2015), such as limited information search strategies, interrupted data analysis and organization, and lower-quality decision making and task performance. In order to overcome these negative consequences, it’s critical to detect cognitive redline promptly and develop effective countermeasures to information overload.

I.2 Solutions to Information Overload

Because information overload is a systemic problem contributed by multiple factors, we should look into its solution from various perspectives by take into account four existing methods: training personnel, activating external automation, redesigning work environment, and redesigning display.

Personal Training, the first method, helps operators to build schema in their long-term memory to automate information retrieval (e.g., LaBerge & Samuels, 1974) and enables operators to adopt efficient information processing strategies (e.g., Kirsh, 2000). The second method external automation can aid human task performance under high workload conditions by actively selecting data, analyzing information, making decisions and selecting actions, and controlling process (e.g., Kaber & Endsley, 2004; Lee & See, 2003; Parasuraman, Sheridan, & Wickens, 2000). The work environments redesign can eliminate redundant or noisy data in the background, therefore reducing the information load (e.g., Chaudhury, Mahmood, & Valente, 2009; Nachreiner, 1995).

The dissertation focuses on the fourth methods: display redesign. As the bridge between human and machine, an ideal display should be able to convey information effectively while requiring less mental efforts for information processing. Following this principle, several display designs were created in the human factors domain, including ecological interface design (EID), multimodal display design, and ambient display design.

I.2.1 Ecological Interface Design

EID is an analytical framework based on *abstraction hierarchy analysis (AH)* and *skills, rules, knowledge taxonomy (SRK).* AH analysis describes the constraints in a work domain into five abstraction levels (i.e., functional purpose, abstract function, general function, physical function, and physical form) and three levels of details (i.e., whole, sub system, and component). SRK describes the mechanisms of human performance into three different levels (Rasmussen, 1999; Vicente, 2002; Vicente & Rasmussen, 1992). They are skill-based behavior (i.e., parallel, automated, and direct information processing and performance), rule-based behavior, (i.e., performance associates with familiar information that regulates human intention and action), and knowledge-based behavior (i.e., serial and analytical problem solving process to cope with unfamiliar and unanticipated situations).

By adopting the results of abstraction hierarchy analysis, EID encourages the use of skill- and rule-based behavior to reduce mental workload and supports knowledge-based behavior for handing the dynamics and novelty of tasks. These characteristics of EID supported it application to the displays of not only small-scale systems, such as neonatal intensive care medicine (Sharp & Helmicki, 1998) and auditory alarm system (Sanderson, Anderson, & Watson, 2000; Watson & Sanderson, 2007), but also large-scale sociotechnical systems, such as the nuclear power plant (Dinadis & Vicente, 1996; Lau, Jamieson, Skraaning Jr, & Burns, 2008).

I.2.2 Multimodal Display Design

A multimodal display transmits information into our brain via multiple sensory channels (e.g., visual, auditory, and tactile channels) so that it can distribute information processing demands among parallel modalities. This property of multimodal display produces several types of benefits on information processing activities, such as synergy (i.e., the merging of information refers to various aspects of the same event or process), redundancy (i.e., the use of more than one modalities to enhance information processing), disambiguation (i.e., the event is clarified by information from different modalities), increased bandwidth of transformed information, and assistance to attention management (e.g., Oviatt, 2002; Sarter, 2002; Sarter, 2006). These cognitive benefits of multimodal display also improve task performance in many work environments. For example, the advanced navigational display that engaged visual and auditory modalities, compared to that only involved visual modality, shortened drivers' reaction time in the navigation task and reduced their errors in the identification task (Liu, 2001). Therefore, multimodal display can be a powerful tool to support cognitive and operational performance under information overload conditions.

In addition, ‘multi-code’ display is modified from multimodal display and can also strengthen human information processing capacity and operational performance by encoding information into multiple dimensions of a display modality, such as spatial location and rhythm pattern of tactile modality (Adroin & Ferris, 2015). A dimension here is defined a sensory feature of modality.

I.2.3 Ambient Display Design

The design of ambient display is inspired by the idea of “calm” technology which is introduced in the ubiquitous computing literature (Weiser & Brown, 1996). An ambient display should be able to ‘quietly’ communicate information to the human at the individual, shared, and public levels (Maclean, 2009). It has been illustrated in a number of interesting examples, such as ambientRoom, which used water ripples, light patches, and natural soundscapes to convey ambient information (Ishii et al., 1998); breakaway, which is a small sculpture that turns into a slouching pose (when people sit for too long) to remind people that they need a stand-up break (Jafarinaimi, Forlizzi, Hurst, & Zimmerman, 2005); a scent diffuser which delivers ambient scent as notifications (Bodnar, Corbett, & Nekrasovski, 2004). Since it fulfills the design goals of distributing information in a manner that minimally loads mental resources, ambient display design may be another promising way to combat information overload problem.

I.3 Evaluation of Display Design

The general goals of all three display redesign can be concluded as 1) improving display informativeness and 2) reducing mental resources required for information processing. However, there haven't been conclusive answers to how we can quantify the extent to which display design achieve the two goals. This part of Chapter I discussed the existing theories and methods that measure display informativeness and required mental resources.

I.3.1 Display Informativeness Measure

Display informativeness reflects the relationship between display and viewer (Woods, 2002) because it is not only affected by display but also constrained by human information processing ability. Since the displayed information can alter the state of each stage of human information processing (i.e., sensation/perception, cognition, and behavior), the level of human response at each stage may be used to illustrate the level of display informativeness.

Sensation/Perception

Display informativeness strongly associates with eye movement pattern (Woods, 2002). This response of visual sensation/perception can be recorded and analyzed by the current eye tracking technologies. For example, the effectiveness of computer display in a visual search task was measured by the length of scanpath and the location and duration of eye fixation (Goldberg & Kotval, 1999). Among these variables of eye movement, the length of scanpath was more sensitive to display effectiveness. It was shorter when the display presentation was well designed but it became longer as the information was poorly organized. Therefore, the eye movement pattern may be a promising indicator of display informativeness at the sensory/perceptual stage. However, in addition to visual response, it's difficult to assess the sensory/perceptual response in other modalities (e.g., auditory and tactile) with the existing technologies.

Cognition

Display informativeness can significantly influence cognitive activities, such as memory recall, monitoring, and decision making. For example, the display that contained more information (i.e., additional functional variables) about the state of a thermal-hydraulic system led to better memory recall of the subset of variables that were most critical to the diagnosis of the current system state (Vicente, 1992). For another example, the less informative displays on the control panel of nuclear power plant (e.g., showing unreliable indicator, failed meter, few emergent features, and unclear reference values) enlarged the difficulty of monitoring the nuclear power plant (Mumaw, Roth, Vicente, & Burns, 2000).

One way to evaluate human cognitive performance during the human-machine interaction is cognitive task analysis (CTA) (Schraagen, Chipman, & Shalin, 2000), which provides complete descriptions of cognitive process and decisions (Clark, Feldon, van

Merriënboer, Yates, & Early, 2008) and better insight into the needs of information from displays (Woods, Wise, & Hanes, 1981). This method has been applied to improve the interface design in complex systems, such as nuclear power plant (Carvalho, dos Santos, Gomes, Borges, & Guerlain, 2008; Woods, Wise, & Hanes, 1981). CTA and other similar analytic tools may play an important role in the evaluation of display informativeness.

Behavior

Finally, the effects of display extend to operators' operational performance (e.g., tracking performance) so that display informativeness may be reflected by behavior measure. For example, the in-vehicle navigational display systems was evaluated by multiple variables of performance, such as average time to input the address of the destination, navigational error, and minimum number of user operations of performing a particular function (Antin, Dingus, Hulse & Wierwille, 1990; Ross & Burnett, 2001).

Among various behavior responses, verbal response plays a key role in the measure of display informativeness. The verbal response (some studies used button pressing to replace verbal response) to external stimuli can be used to calculate *information transmitted* (i.e., the amount of information that is successfully transmitted from displays to human observers regarding the lost information during the transition), which is superior to other indicators in terms of its capacity to illustrate the quantity of displayed information. In the calculation of information transmitted, the verbal response provides input (i.e., event probability) into the mathematical formula (built upon the concept *entropy*) to quantify the discrete information transmitted (Miller, 1954; Shannon, 1948), which has been successfully used to quantify the information displayed via different sensory channels, such as visual (e.g., Hsia, 1971), auditory (e.g., Hsia, 1971; Pollack, 1952; Pollack, 1953), and tactile (e.g., Tan, Reed, & Durlach, 2010). However, the event probability in the calculation can be only obtained in carefully-controlled experimental environments so that it's difficult to be applied to the real-world environments. Moreover, it's difficult to extend the quantitative measure of discrete information to continuous and mixed (discrete & continuous) information in the current experimental settings.

I.3.2 Mental Workload Measure

The information processing activities impose mental workload on our cognitive systems, as an analogy to physical workload. Mental workload is a complex concept with various characteristics (Xie & Salvendy, 2000) which are listed below.

- Mental workload cannot be detected directly.
- Mental workload can be static or dynamics over a period of time.
- Mental workload is a multi-dimensional variable because our brain reacts to the perceived information in many different ways, such as memorizing, calculating, and reasoning.

- Mental workload can be affected by multiple human factors, either long-term factors such as cognitive capacity or short-term factors such as emotion and fatigue.
- According to Cognitive Load Theory, mental workload consists of three types of workloads: intrinsic, extraneous, and germane (Paas, Tuovinen, Tabbers & van Gerven, 2003).

There are four major categories of mental workload measures: performance measure, subjective self-reported measure, physiological measure, and mental modeling. In the early years, mental workload measures relied on the evaluation of primary and secondary task performance, eventually being extended to subjective self-reported measures. Along with the development of biotechnologies and computational technologies, physiological measures are becoming the primary tools to measure mental workload and mental modeling.

Performance Measure

Primary- and secondary- task performances (e.g., choice reaction time, memory, monitoring, and tracking) were used to indicate mental workload (Ogden, Levine, & Eisner, 1979).

Subjective Self-reported Measure

This type of measure is based on questionnaires that collect participants' subjective ratings and weightings to selected scales, such as NASA Task Load Index (NASA TLX; Hart & Starvland, 1988), Subjective Workload Assessment Technique (SWAT; Reid & Eggemeier, 1982), and Workload Profile (WP; Tsang & Velazquez, 1996).

Physiological Measure

Physiological measure relies on the technologies that continuously collect physiological data which strongly correlate with mental workload, such as hear rate variability, skin conductance response, and pupil diameter.

Mental Modeling

The modeling techniques simulate the cognitive activities of information processing and assign a 'workload score' to each activity. The overall mental workload is indicated by the sum of the scores. The well-known mental modeling techniques include adaptive control of thought – rational (ACT-R; Anderson, Matessa, & Lebiere, 1997), queueing network-model human processor (QN-MHP; Liu, Feyen, & Tsimhoni, 2006), and Visual/Auditory/Cognitive/Psychomotor (VACP; Aldrich, Szabo, & Bierbaum, 1989; McCracken & Aldrich, 1984).

Although various mental workload measures are developed by researchers in different fields, there are three challenges that haven't been overcome to date. First, no single measure was found to be sensitive enough to mental states under all kinds of task scenarios (Matthews, Reinerman-Jones, Barber, & Abich, 2015; Mehler, Reimer, & Coughlin, 2009). Since mental workload is a multi-dimensional variable, a single measure may be sensitive to some of its dimensions but less sensitive to others (Matthews, Reinerman-Jones, Barber, & Abich, 2015). To build a robust mental workload measure, several studies proposed various ways to algorithmically combine multiple mental workload measures (Miyake, 2001; Ryu & Myung, 2005; Tan, Reimer, Mehler, & Coughlin, 2011), showing a promising improvement on the accuracy of measures.

Secondly, it's difficult to detect mental overload in a timely manner. Mental overload happens when mental workload reaches cognitive redline. Thus, the detection of cognitive redline is critical to stop mental overload. The existing detection methods rely on the degradation of task performance and high self-reported ratings of mental workload, but they could be less reliable or sensitive. However, some studies showed that the upper or lower limit of the arousal of sympathetic nerve systems may strongly associate with cognitive redline, such as the maximum pupil diameter (Juris & Velden, 1977), the minimal prefrontal activation reported by fNIRS (Durantin, Gagnon, Tremblay, & Dehais, 2014), and the lowest heart rate variability (Rodriguez Paras, Yang, Tippey, & Ferris, 2015). These findings suggest the physiological patterns as useful indicators of cognitive reline, but the idea needs further validation.

I.3.3 Cognitive Efficiency Measure

The above measures of either display informativeness or imposed mental workload may not provide sufficient insight into display impacts. For example, a display that conveys larger amount of information is not necessarily the better display because it may also impose higher mental workload on human operators. Therefore, it is necessary to evaluate a display regarding its effectiveness and cost.

Efficiency is a proper display property that can take into account both display effectiveness and cost. According to the Processing Efficiency Theory, *processing efficiency* (similar to display efficiency) is defined as the ratio of performance effectiveness to mental effort, in which performance effectiveness is denoted by the quality of performance (Eysenck & Calvo, 1992; Eysenck, Derakshan, Santos, & Calvo, 2007). However, the relevant studies haven't proposed a quantitative way to measure each aspect of processing efficiency. The goal of these studies was to understand the anxiety effect on processing efficiency and found that anxiety increased working memory load (e.g., Fales, et al., 2008), reduced mental storage and impaired processing efficiency (Eysenck & Calvo, 1992).

In addition to processing efficiency, the similar concept of cognitive efficiency has been studied in many other domains, such as graphic design (Carner & Larking, 1989) and educational psychology (Clark, Nguyen, & Sweller, 2011) , but it is usually indicated by how fast people complete the tasks. For example, the cognitive efficiency in the processing of graphics was indicated by the response time (Carner & Larking, 1989). The efficiency of visual processing in driving simulation task was denoted by the search rate or response time (Murray & Jannelle, 2003; Wilson, Smith, Chattington, Ford, & Marple-Horvat, 2006). However, response speed, as a single-scale indicator, provides limited insight into each component of cognitive efficiency.

Finally, the research on *instructional efficiency* (i.e., efficiency of instructional design) provided the measure of each aspect of efficiency: task performance (as instructional effect) and cognitive load (as instructional cost) (Paas, Tuovinen, Tabbers & van Gerven, 2003; Paas & van Merriënboer, 1993). In the computation of instructional efficiency, task performance is indicated by the exam score and cognitive load is measured by a single-scale self-reported rating. The exam score and self-reported rating can be combined in two ways. In the first way, the two types of measures generated the x-y dimensions of a two-dimensional Euclidean plane. Instructional efficiency was indicated by the perpendicular distance from a point on the plane to the diagonal which represents zero efficiency (e.g., Paas, Touvinen, Tabbers, &Van Gerven, 2003; Paas & van Merriënboer, 1993). In the second combination, instructional efficiency was calculated as the ratio of task performance to cognitive load (Kalyuga & Sweller, 2005).

The studies on instructional efficiency was the first effort to develop the quantitative evaluation of both dimensions of efficiency. The measures of instructional efficiency have been validated in multiple learning scenarios, including example-based training, older learners' instruction, multimedia learning, and computer-supported collaborative learning (Kirschner, 2002). Similar methods have also been used to evaluate the efficiencies of various training programs (Fiore, Scielzo, Jentsch, & Howard, 2006; Salden, Paas, Broers, & van Merriënboer, 2004) and information portals such as websites (lo Storto, 2014).

I.4 Gap of Knowledge

The previous studies only evaluated the cognitive efficiency of instructional materials (e.g., Paas, Touvinen, Tabbers, &Van Gerven, 2003), which mainly engaged learning activities, or e-commence websites, which required limited motoric response (e.g., lo Sortor, 2013). This dissertation, however, aims to evaluate displays that support real-time operational task in human-machine systems, which is under more complicated context.

In addition to using different context, the dissertation also aims to address the limitations of current cognitive efficiency measures. First, the existing measures used task performance to indicate display effectiveness, but task performance might not be sensitive enough to display informativeness because it also largely affected other factors, such as individual differences in motor skill. Moreover, performance measures, such as task-complete time, response accuracy, and test score, cannot illustrate the quantity of displayed information which is a critical aspect of quantitative evaluation of display.

Second, the single-dimension rating scale used in instructional efficiency measure may not be sensitive enough to mental workload in multitasking condition. Also, the subjective measure cannot provide continuous and high-resolution monitoring of mental workload during the task.

Third, the existing cognitive efficiency measures did not consider the effects of relevant contextual factors on them.

I.5 Research Questions and Contribution

To fill the gap of knowledge, the dissertation aims to answer five research questions.

I.5.1 How Can We Measure Each Dimension of Cognitive Efficiency of Display?

Different from instructional display or ecommerce website, the displays studied here are those which support operational performance in human-machine systems. Given that cognitive efficiency consists of two dimensions (display informativeness and imposed mental workload), the first question aims to determine the effective measures of each dimension. This question will be answered by the examples which apply the measures of display informativeness or mental workload and provide insight into the advantages and limitations of each measure.

I.5.2 How Can We Measure the Cognitive Efficiency of Basic Displays in Single-Task Condition?

In this question, the basic displays are a visual display that present RGB colors and an auditory display that delivers pure tones. The colors and pure tones vary in intensity (brightness vs. loudness) or spectrum (hue vs. pitch). This question aims to validate the measure of displays' cognitive efficiency under a low-workload condition. Moreover, the study for this question can illustrate the differences in cognitive efficiency between modalities or dimensions.

I.5.3 How Can We Construct the Cognitive Efficiency Measure That Is the Most Predictive of An Operator's Multitask Performance?

We hypothesized that display that shows higher cognitive efficiency is able to improve the operator's multitask performance in human-machine systems. The study aims to demonstrate this hypothesis by showing the positive correlation between cognitive efficiency of display and multitask performance, which provides theoretical support for the adoption of cognitive efficiency measure. Moreover, since each dimension of cognitive efficiency can be measured in many ways, it's necessary to know which of these measures can be combined as the cognitive efficiency measure that has the highest predictive power of multitask performance.

I.5.4 How Can We Measure the Cognitive Efficiency of Novel Speedometer Displays in Multitask Conditions?

The novel speedometer displays are different from the displays studied in previous questions. First, the novel speedometer displays present continuous information instead of discrete information. Second, the novel speedometer displays engage different perceptual modalities, such as peripheral-visual, auditory, and tactile modalities, compared to the previous ones which only engaged focal vision. Third, the auditory and tactile displays engage the perception of more complex dimensions, such as beat pattern. The findings of this question will expand our understanding of cognitive efficiency of more diverse displays.

I.5.5 How Do Novel Speedometer Displays Affect Concurrent Tracking Performance in Driving?

For this question, we will explore the function of novel speedometer display on supporting drivers' concurrent tracking performance. The findings of this study can provide valuable guidance for the display design for multitask performance under high workload condition. The study will also examine relationship between cognitive efficiency of novel speedometer display and concurrent tracking performance to see if it is consistent with the findings of Q3.

I.6 Dissertation Structure

To answer these research questions, the dissertation covers my eight studies completed between 2011 and 2016. These studies are organized into three sections.

I.6.1 Measurement Techniques for Cognitive Efficiency (Chapter II - III)

In the first section, Chapter II answered Q2 by using examples to illustrate the measures of display informativeness and mental workload. This chapter also discussed the detection of cognitive redline because cognitive efficiency measure is especially important in high mental workload condition. Based on the measures in Chapter II, Chapter III proposed Cognitive Efficiency metric and initially validated it with the simple displays.

I.6.2 Cognitive Efficiency and Multitask Performance (Chapter IV)

The second section only includes Chapter IV which linked cognitive efficiency to multitask performance - an important topic in human factors domain - and examined the critical hypothesis in Q3. Moreover, this chapter determined the version of CE metric that was the most predictive of multitask performance.

I.6.3 Cognitive Efficiency of Novel Speedometer Display (Chapter V - VII)

Before moving into the study of novel speedometer display, Chapter V investigated the perception of beat pattern, which was one of the display's special dimensions. In Chapter VI, five novel speedometer displays were evaluated in terms of their informativeness, imposed mental workload, and cognitive efficiency. Chapter VII, the last chapter, applied novel speedometer displays in a practical driving scenario: concurrent tracking task.

CHAPTER II

MEASURING THE DIMENSIONS OF COGNITIVE EFFICIENCY OF DISPLAY*

According to the literature review in Chapter I, display informativeness and imposed mental workload are the two critical dimensions of cognitive efficiency. This chapter answered the first research question (Q1) in Chapter I: **how can we measure each dimension of cognitive efficiency of display**. The first section of this chapter illustrated how to mathematically quantify display informativeness based on Information Theory. In the second section, we described a case study of mental workload measure (based on Toyota Economics Settlement Safety Research), which used two physiological measures - heart rate variability (HRV) and skin conductance response (SCR) - to evaluate drivers' cognitive states under short-term loads and acute stress events. In the end, this chapter discussed how to use physiological measures to detect cognitive redline, which is one of the challenging topics for mental workload measure.

* Part of this chapter is reprinted with permission from *Proceedings of the Human Factors Society Annual Meeting*, 60, 1.

II.1 Calculation of Information Transmitted

The information transmitted, which is an important indicator of display informativeness, can be quantified into bits based on the mathematical theory of communication (Information Theory; Shannon, 1948). The theory is one of the most important scientific theories in the 20 century. Information Theory has deeply influenced a wide range of domains, including electrical engineering, computer science, physics, philosophy, and economics.

In the theory, the communication system consists of five parts: information source, transmitter, channel, receiver, and destination (See Figure 1). The information travelling (arrows in Figure 1) in the communication system can be roughly classified into three main categories: discrete, continuous and mixed (discrete & continuous). The study here focuses on the quantification of discrete information because it's difficult to quantify the continuous and mixed information under the current experimental settings.

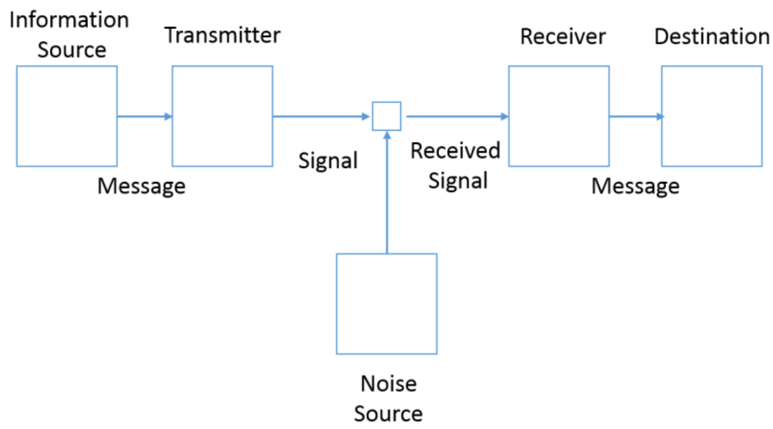


Figure 1. The framework of communication system.

The discrete information consists of a sequence of discrete symbols (e.g., a sequence of letters) which appear in certain (conditional) probabilities. The probabilities of discrete symbols (p_i, i represents each symbol) need to be input into *entropy* (i.e., $H = -\sum p_i \, log \, p_i$) to calculate the amount of discrete information.

Based on the event probability and the concept of *entropy*, Miller (1952) showed a way to calculate information transmitted in the experimental settings. The example here explains the way of calculating information transmitted (*H*). Here are two types of stimuli (A and B) and each one presents totally 8 times to observes. Observes need to verbally response to each stimuli (or press corresponding buttons). The number of each S-R pair is listed in Table 1, which enables us calculate the joint probability of each S-R pair.

Table 1: Stimulus-Response Confusion Matrix

	Response A	Response B	
Stimulus A	6	2	8
Stimulus B	1	7	8
	7	9	

The joint probability will be input into the equation $H = -\sum p_i \, log \, p_i$ to calculate the joint entropy (*H(R, S)* in Equation 3). In addition to conditional entropy, we also need to obtain the marginal probability of each response and stimuli, which is used to calculate the entropy of response (*H(S)* in Equation 1) and stimuli (*H(R)* in Equation 2). In the end, the information transmitted (mutual information; I (S: R) in Equation 4) is calculated as entropy to minus condition entropy.

$$H(S) = -\left(\frac{8}{16}\, log_2\left(\frac{8}{16}\right) + \frac{8}{16}\, log_2\left(\frac{8}{16}\right)\right) = 1 \; bits \tag{1}$$

$$H(R) = -\left(\frac{7}{16}\, log_2\left(\frac{7}{16}\right) + \frac{9}{16}\, log_2\left(\frac{9}{16}\right)\right) = 0.99 \; bits \tag{2}$$

$$H(R, S) = -\frac{6}{16}\, log_2\left(\frac{6}{16}\right) - \frac{2}{16}\, log_2\left(\frac{2}{16}\right) - \frac{1}{16}\, log_2\left(\frac{1}{16}\right) - \frac{7}{16}\, log_2\left(\frac{7}{16}\right) = 1.68 \; bits \tag{3}$$

$$I\,(S{:}R) = H(S) + H(R) - H(R, S) = 1 + 0.99 - 1.68 = 0.31 \; bits \tag{4}$$

II.2 Evaluating Cognitive States under Short-term Loads and Acute Stress Events - A Case Study of Mental Workload Measure

According to Chapter I, mental workload can be measured by numerous methods from several categories: subjective self-reported measure, performance measure, physiological measure, and mental modeling. A number of popular mental workload measures are summarized in Appendix 1 and described in terms of their measured variables, advantages, and limitations. These measures have been applied for mental workload management in a wide range of work environments, such as vehicle cockpit (De Waard, 1996), aircraft cockpit (e.g., Wilson, 2002), air traffic control station (e.g., Brookings, Wilson, & Swain, 1996), nuclear power plant (e.g., Jou, Yenn, Lin, Yang, & Chiang, 2009), surgery room (e.g., Carswell, Clarke, & Seales, 2005), and instructional learning environments (e.g., Wiebe, Roberts, & Behrend, 2010).

As a case study, this section focused on two physiological measures - HRV and SCR – and described their applications in Toyota Economic Settlement Safety Research during the 2015 calendar year. HRV and SCR were applied to provide online detection of arousal levels of a driver's sympathetic nervous system, which associated with the potentially-problematic cognitive states that are detrimental to safe driving behavior. The cognitive states of interest in this section were those under three types of short-term loads (i.e., mental, emotional, and motoric) and one acute stress event (i.e., unintended acceleration).

II.2.1 The Impact of Cognitive Loads on Physiological Arousal

Exposure to cognitive or physical stressors will increase arousal of the human autonomic nervous system, which initiates the "fight or flight" response in the body and mind. Heightened sympathetic arousal can have a number of effects that are at times beneficial and at other times detrimental to the safety of the human. This heightened arousal generally tends to increase one's spatial awareness of surroundings and prepare the human for an imminent need to move quickly. It also gives room to in-attentional blindness; important events are ignored since they seem unimportant to the goal.

The sympathetically-aroused mental state can be detrimental to a modern human in control of complex systems such as a vehicle. Having identified 'heightened state of sympathetic arousal' as a potentially-problematic driver state, physiological indicators can be used to detect this state of arousal online and used to infer a problematic driver state. Identifying this state can then trigger an automated mitigation strategy to better support the driver. Because sympathetic arousal causes several physiological changes in the body, such as in perspiration and cardiac dynamics, physiological indicators may be used to measure the impact of cognitive stressors. Notably, stressors are found to strongly influence heart rate variability (HRV; Taelman, Vandeput, Spaopon, & Van Huffel, 2009) and skin conductance response (SCR; Boucsein, 2012; Shi, Ruiz, Taib, Choi, & Chen, 2007; Villarejo, Zapirain, & Zorrilla, 2012).

Heart Rate Variability

Heart rate variability (HRV) is the variance of beat-to-beat heart interval in a time window. By identifying corresponding points in electrocardiogram (ECG) signatures (such as the R-R interval) and measuring the intervals between these points in successive heartbeats observable in an electrocardiaogram as shown in Figure 2, the variance among the intervals can be calculated over a specified time window to represent the measure of HRV (Karim, Hasan, & Syed, 2011).. As cognitive stress levels increase, HRV tends to decrease; heartbeats become more regular, less variant. However, stressors can also cause the underlying heart rate to increase (another sympathetic nervous system response), which also corresponds with a substantial increase in HRV when the time window sampled is narrow. For this reason, HRV is generally used to determine differences in sympathetic arousal over larger windows of time (Task Force of the European Society of Cardiology, 1996).

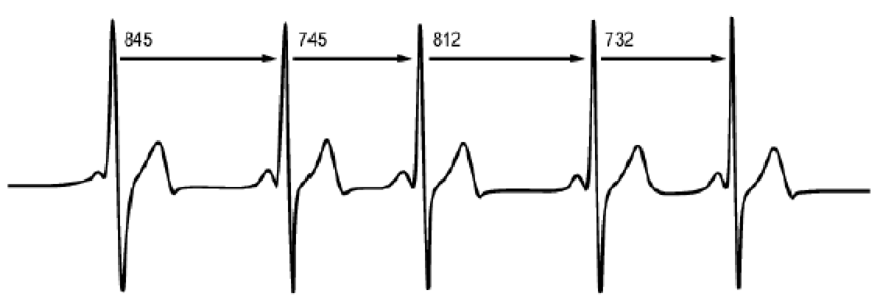


Figure 2. Sample ECG reading, with R-R intervals varying from 732 ms to 845 ms. Image source: Heart Rate Variability (HRV) | Polar USA (2016).

Skin Conductance Response

Skin conductance response, or "galvanic skin response", is a phenomenon that occurs when sympathetic arousal causes the surface of the skin becomes a better conductor of electricity due to increased perspiration in the skin (Boucsein, 2012). Since skin perspiration is a function of several biological and environmental factors (including body and skin temperature, airflow, and humidity of the surrounding environment), absolute levels of skin conductance (referred to as Skin Conductance Level, SCL) are less informative than observations of rapid changes in skin conductance over short time windows. These are indicative of increased sympathetic arousal that follows from perception of a stress-inducing event or context. The changes in sympathetic arousal can be inferred by the presence of “responses”, the amplitude of those responses, and the frequency of their occurrence.

Skin conductance response can be defined and analyzed in a number of different ways. As a first step, “responses” must be identified in the raw data. Responses can be identified as characteristic increases in conductance following a stimulus presentation, with the threshold values used to identify the occurrence of a response. Figure 3 below illustrates a generic response pattern in skin conductance data (Kappeler-Setz, Gravenhorst, Schumm, Arnrich, & Tröster, 2013).

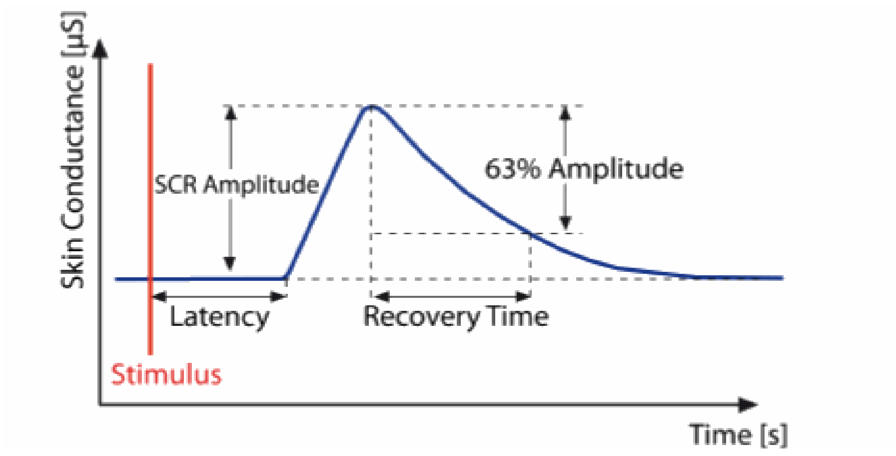


Figure 3. A characteristic pattern in skin conductance that constitutes a skin conductance "response". Measures of interest include SCR Amplitude, as well as the count or frequency of these responses over a predefined window of time (image from Kappeler-Setz, et al., 2013).

In the current study, we focus on two measures of skin conductance: SCR-Amplitude and SCR-Frequency. After identifying each "response", the amplitude of that response can be calculated as the change in skin conductance level observable over the duration of the response, which will be specified in the current research as SCR-Amplitude. In contrast, SCR-Frequency is the count or calculated frequency of responses identified within a predetermined time window.

Physiological changes occur following the perception of an event or context that activates the "fight or flight" response. HRV is measured over a window of time that, by definition, must include several heartbeats while SCR is limited in resolution primarily by the threshold of response, the different dynamics in the observed patterns in these two measures may provide additional levels of sensitivity and specificity in identifying the several potentially problematic cognitive states under which the driver may be operating.

The efforts described in this paper focus on physiological data collected via relatively inexpensive, lightweight sensors that could be positioned in common driver contact points in a vehicle, such as in the steering wheel or seatbelt. By comparing and contrasting the results among these sources of driver data, a clearer set of conclusions can be drawn about an effective suite of human sensors that can support detection of problematic driver cognitive states.

II.2.2 Method

There were 88 participants in the study with over two data collection periods. The demographics of participant are illustrated in Table 2.

Table 2: Participant Demographics of the Simulation Study

	Older(>65) drivers	Younger(<25) drivers	Total
Males	26	20	46
Females	21	21	42
(Total)	47	41	88

Apparatus

The CTS simulator facility in Texas Transportation Institute supports a medium-fidelity driving simulation environment consisting of three large high-resolution monitors, bucket seat, and realistic steering wheel and pedal system (See Figure 4). Driving scenario consisted of a simulated rural highway with varied loaded segments.

Figure 4. Driving simulation environments in Toyota Economics Settlement Safety Research.

The projects applied multiple metrics to evaluate human and vehicle states. However, because of the volume of a proceeding paper, we only focus on two types of physiological measurement techniques: Zephyr Bioharness3 (BioHarness 3 Zephyr Technology Corporation, 2015)., a lower-chest strap with microcontroller and sensors embedded on it to measure HRV and Shimmer3 (Shimmer3 GSR Unit, 2016) consisted of a module attached on the wrist band and two sensors on the palm of monodominant hand to measure SCR.

Tasks

The resting baseline of all physiological variables were were collected in the "baseline" condition in which participants only needed to relax and listen to soothing music for five minutes. Then, they were familiarized with the driving condition in the "practice drive". After practice, they were asked to complete 6 driving conditions (Details see Table 3). Before each driving condition, participants were asked whether they were experiencing simulator sickness. The experiment would stop if the answer was "yes".

The effects of four independent variables were analyzed for each dependent measure:

- *Driving condition*, a within-subject factor with 6 levels (See Table 3);
- *Age*: young (age < 25) vs. old (age > 65), as a between-subjects factor;
- *Sex:* male vs. female, as a between-subjects factor;
- *Failure loading*: no loading vs. full (cognitive + emotional + motoric) loading in the F driving condition, as a between-subjects factor.

Table 3: The Description of each Driving Condition in the Simulation Study

Driving Condition	Description
Normal	Participants drove the length of the experimental scenario, but no additional loading was present for the entirety of the drive
LD1~LD4 Balanced	4 Loaded drives (LD): Experimental drives that increased the driver loading by introducing high-activity construction zone areas that heightened the load on the driving task, and additional secondary task loads:
LD1-driving	Construction zone loading only, **no secondary task** loading
LD2-cognitive	Secondary cognitive loading via experimenters asking challenging ***analytical and mathematical questions***
LD3-emotional	Secondary emotional loading via experimenters asking **emotionally-charged questions**
LD4-motoric	Secondary motoric loading via a **texting task** completed on a smartphone
Failure-event condition	"Failure" drive which either replicated the conditions of LD1 or included cognitive, emotional, and motoric loading tasks throughout the drive. At the completion of the drive, an unexpected Failure-Event (unintended acceleration) is triggered. This occurs while the participant is waiting at a red light and a car pulls up in front of them. Then, suddenly, the vehicle accelerates and the participant has 10 second to avoid crashing into the car in front. Since their brake also fails, they should steer to the sides to avoid a collision.

In each of these scenarios, two 2.5 minute intervals are secretly defined to collect higher cognitive loading states. Those 2.5 minute intervals in each scenario, start and end based on designed marker locations on the road, which provided us the ability to have reference windows from each scenario to compare the data in between-scenarios.

Procedure

After signing the consent form and completing pre-study survey for background and mental states, participants were introduced to driving simulator and scenarios. Then, they needed to wear Bioharness3 on their lower chest and Shimmer2 on the wrist of their non-dominant hand (see details via the links in method section). Then, they were asked to adjust the height of seat and steering wheel to their comfort. Prior to the start of the experiment, all the devices were calibrated and recordings initiated. Participants completed baseline condition first followed by the practice condition. After practice, participants continued to complete the 6 driving conditions, among which the loaded driving conditions (LD1 ~ LD4) were balanced with the Failure-event condition always finishing last. In the end, participants were asked to complete the post-study surveys and compensated $40.

Data Analysis

The proceeding paper reported the analytical results of three dependent variables: HRV, SCR-amplitude, and SCR-frequency. Three-way (driving condition × age × sex) ANOVAs were used to analyze the dependent variables in driving conditions. In the Failure-event condition, four-way (driving condition × age × sex × loading) ANOVAs were used to analyze the dependent variables. Post hoc tests used multiple comparison with Bonferroni correction. All analyses were complete in R 3.1.3.

II.2.3 Results

Only complete data sets obtained from participants were analyzed in the study. Only the most important findings of each dependent variable were reported in this section.

HRV

Mean HRV across All Driving Conditions. For the analysis of HRV, 61 participants provided complete HRV datasets that could be entered into the statistical model (older males: N = 16; older females: N = 14; younger males: N = 16; and younger females: N = 15).

The main effects of *driving condition* ($F(3, 341)=11.99$, $p<.001$, $\eta^2=0.039$) and *age* ($F(3, 341)=338.83$, $p<.001$, $\eta^2=0.218$) significantly affected the measures of HRV. Figure 5 below illustrates the mean HRV values across the entirety of each driving condition and the significant results of post hoc test can be found in Table 4. There was no significant interaction effect found among any of the independent variables.

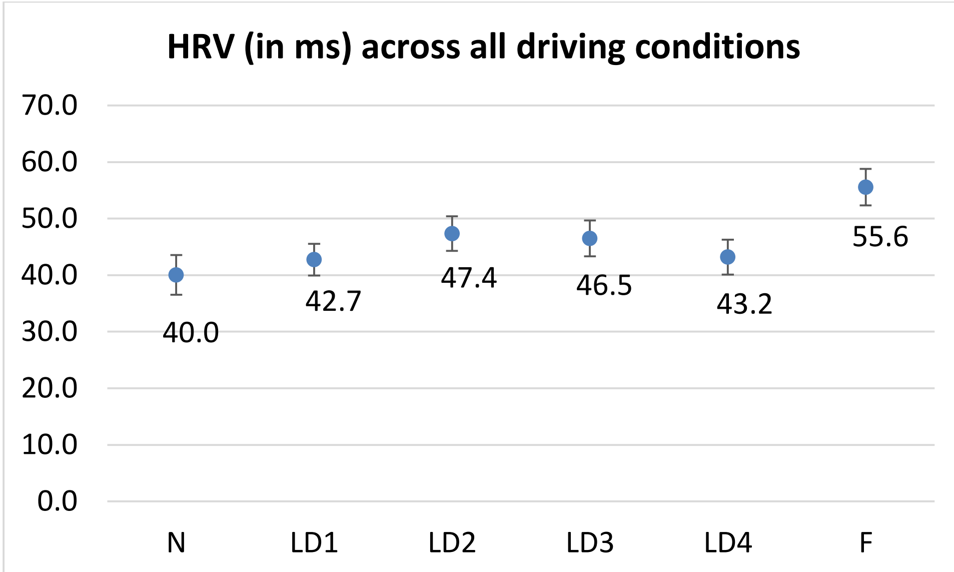


Figure 5. Mean HRV across all driving conditions.

Table 4: Post-hoc Results of Driving Condition and Age on HRV

Significant main effect	Post Hoc Tests (Significant Results)
driving condition	F (55.6 ms) > N (40.0 ms) F (55.6) > LD1 (42.7) F (55.6) > LD2 (47.4) F (55.6) > LD3 (46.5) F (55.6) > LD4 (43.2) N < LD2 (47.4) N < LD3 (46.5) N < LD4 (43.2) LD1 (42.7) < LD2(47.4) LD1 (42.7) < LD3 (46.5)
age	Younger (57.5 ms) > Older (33.9 ms)

Mean HRV in Loaded Driving Conditions. For this analysis, 54 participants provided complete HRV datasets that could be entered into the statistical model (older males: N = 14; older females: N = 14; younger males: N = 12; and younger females: N = 14).

Within the loaded driving conditions only (LD1, LD2, LD3, and LD4), significant effects were found for *driving condition* (F(3, 999)=11.90, p<.001, η^2=0.007), *segment* (F(4, 999)=33.31, p<.001, η^2=0.027), *age* (F(3, 999)=1049.7, p<.001, η^2=0.216) and *sex* (F(1, 999)=4.58, p<.001, η^2=0.033) on HRV. Figure 6 below illustrates the mean HRV values in each segment of each loaded condition, and the significant results of post hoc test can be found in Table 5.

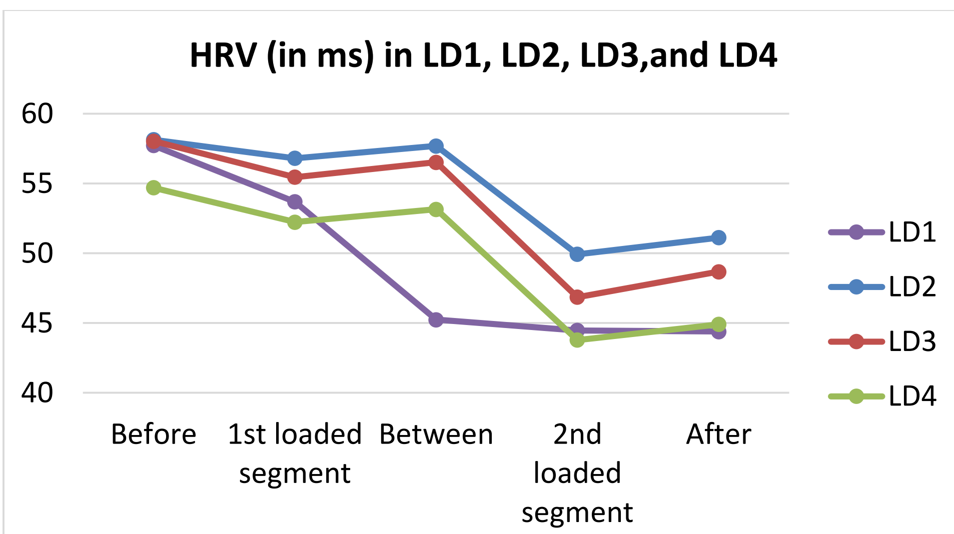


Figure 6. The mean HRV under each segment of each Loaded driving condition. The loaded segments S2 and S4 are affected by the secondary tasks in the driving conditions.

Table 5: Post-hoc Results of Significant Main Effects on HRV in Loaded Conditions

Significant main effect	Post Hoc Tests (Significant Results)
driving condition	LD1 (42.3 ms) < LD2 (48.1 ms) LD1 (42.3) < LD3 (46.8) LD2 (48.1) > LD4 (44.3) LD3 (46.8) > LD4 (44.3)
segment	S1 (51.2 ms) > S2 (48.1 ms) S2 (48.1) > S3 (46.7) S3 (46.7) > S4 (39.9)
age	Younger (60.0 ms) > Older (33.7 ms)
sex	Males (46.2 ms) > Females (44.6 ms)

SCR-Amplitude

SCR-Amplitude across All Driving Conditions. While no significant main effects were identified, a significant interaction effect between *age* and *sex* was found ($F(1,204)=10.05$, $p<.001$), as illustrated in Figure 7. Post hoc test showed that younger males showed higher SCR-Amplitude than did the older males ($p < .05$), but no difference was found between young and old females. Younger males showed higher SCR-Amplitude than younger female participants ($p < .05$).

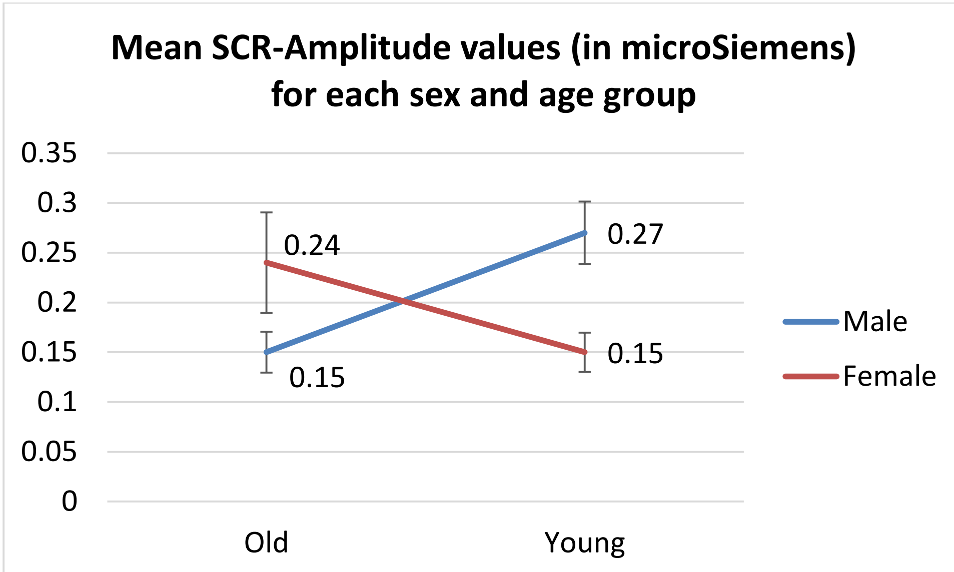


Figure 7. The mean amplitude of SCR interaction effect between sex and age.

SCR-Amplitude in the Failure-Event Condition. For this analysis, 44 participants provided complete SCR-Amplitude datasets that could be entered into the statistical model (older males: N = 11; older females: N = 12; younger males: N = 8; and younger females: N = 18).

The influence of the Unintended Acceleration (UA) event lasted for approximately 5 seconds. Therefore, we compared the mean amplitude of SCR within 5 second windows before and after UA. The results showed the UA occurrence significantly affected SCR-Amplitude ($F(1,66)=5.16$, $p<.002$). The mean SCR after UA (SCR-Amplitude = 0.34 microSiemens) was significantly larger than the mean before UA (0.08 microSiemens). See Figure 8.

Interestingly, this includes roughly equivalent contributions from participants in both the no loading and full loading groups according to the *failure loading* between-subjects variable, suggesting a robust way to detect sympathetic arousal in response to unexpected and sudden events such as UA.

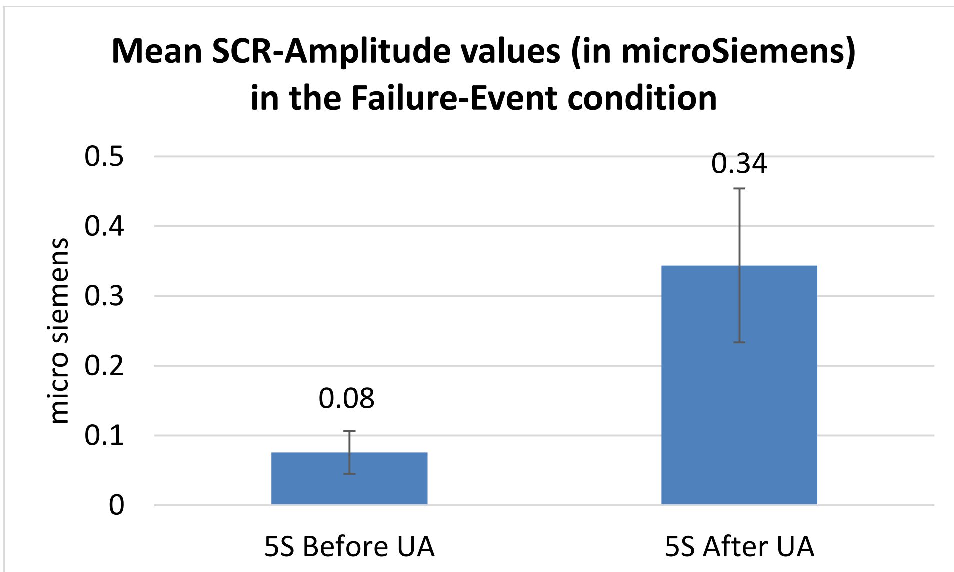


Figure 8. The mean amplitude of SCR immediately before and after the unintended acceleration event.

SCR-Frequency

Frequency of Skin Conductance Response. The frequency of responses (SCR-Frequency) was calculated as the number of responses per minute and were analyzed to compare across all driving conditions, within the defined segments of each loaded condition (LD1, LD2, LD3, and LD4), and in comparing fully-loaded and unloaded participants during the failure drive.

SCR-Frequency across All Driving Conditions. For this analysis, 38 participants provided complete SCR-Frequency datasets that could be entered into the statistical model (older males: N = 13; older females: N = 8; younger males: N = 7; and younger females: N = 10).

A significant main effect of *age* was found for SCR-Frequency across all driving conditions ($F(5, 203)=2.72$, $p=.021$). Older participants (mean = 7.7 responses/min) showed higher frequencies than did younger participants (3.9 responses/min).

SCR-Frequency in the Loaded Driving Conditions. For this analysis, 37 participants provided complete SCR-Frequency datasets that could be entered into the statistical model (older males: N = 13; older females: N = 7; younger males: N = 6; and younger females: N = 11).

Significant main effects of *driving condition* ($F(5, 659)=3.46$, $p=.016$), *segment* ($F(4, 659)=2.40$, $p=.049$), and *age* ($F(1, 659)=31.85$, $p<.001$) were found. Figure 9 illustrates the pattern of results across each loaded condition (LD1 - LD4) and within each segment. Significant results of post hoc test can be found in Table 6.

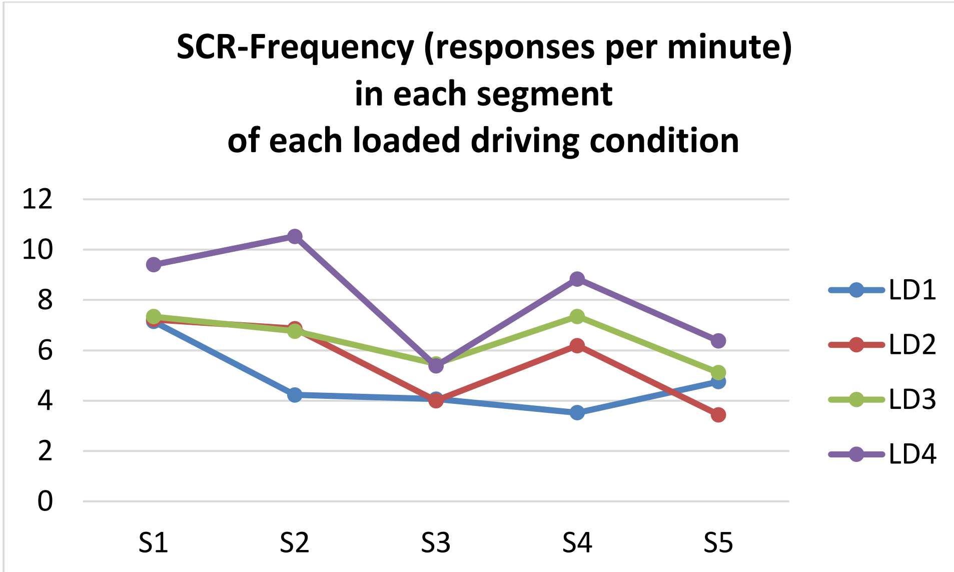


Figure 9. Mean SCR-Frequency values in each segment of each loaded driving condition.

Table 6: Post-hoc Results for Significant Main Effects on SCR-Frequency

Significant main effect	Post Hoc Tests (Significant Results)
Driving condition	LD1 (4.75 responses/min) < LD3 (6.41) LD3 (6.41) < LD4 (8.10) LD2 (5.55) < LD4 (8.10)
Segment	S1 (7.78) > S3 (4.72) S1 (7.78) > S5 (4.92) S2 (7.78) > S3 (4.72) S2 (7.78) > S5 (4.92) S4 (6.48) > S3 (4.72) S4 (6.48) > S5 (4.92)
Significant main effect	Post Hoc Tests (Significant Results)
Age	Older (8.2) > Younger (3.8)

II.2.4 Discussion

Age and Sex Differences

Several analyses showed significant effects of sex and/or age grouping on the indicators of sympathetic arousal. These should be strongly considered. Physiological responses seem to change as a function of age, and presumably sex as well. But perhaps more interestingly, members of each group may experience differing levels of sympathetic arousal in response to loading conditions and events. For example, younger drivers are likely better-practiced in the task of texting on a smartphone, and perhaps also in the ill-advised practice of texting while driving. Therefore, it could be expected that the sympathetic arousal directly attributable to the secondary texting task may have been less dramatic for younger drivers. Similarly, emotionally-charged questions may elicit stronger responses for one of the two sexes, depending on the nature of the questions and whether the questions are being asked by an experimenter who is of the same or the opposite sex.

In analyses across all driving conditions, there is a clear trend that shows arousal increases from the lowest-loading normal drive "N" condition, through the 4 loaded conditions, and shows the highest arousal in the "F" failure-event condition. Whether or not they reached significance (did with the HRV measures), a clear increase in sympathetic arousal indicators show that a sudden, unexpected event, such as the UA failure event can be reliably detected. This is the case regardless of whether the failure drive was completed under no loading or full loading contexts.

Pending further analysis, it appears that HRV measures are more sensitive than SCR measures in detecting and differentiating the types of loading. LD1, which is a loaded drive but does not include a secondary task (such as answering questions or texting) can be reliably distinguished from other loaded conditions with SCR, however LD2 – LD4 are not reliably distinguished. It is useful to be able to differentiate driving-induced loading from (secondary task + driving)-induced loading, and HRV offers a promising means to do this.

Detecting Changes in Loading

Both HRV and SCR-Frequency showed an ability to significantly detect changes in loading, when observed across segments in loaded conditions. Interestingly, while the loaded conditions S2 and S4 could be distinguished from the other conditions, the pattern of changes among segments differed between these two measures.

While HRV shows a general decrease as the scenario unfolds (illustrating the effects of loading over time), SCR-Frequency is relatively stable when viewed across the scenario. This illustrates that HRV may be better for indicating longer-term changes in loading, on the order of several minutes, while SCR may be less sensitive to analyses over longer time windows. HRV also does not appear to "recover" to illustrate a decrease in arousal that should be expected after loading task segments are completed. With one curious exception in LD1, HRV does not change very much from S1 – S3, but the second loading segment in S4 shows a precipitous drop in all loading conditions, illustrating an increase in workload/stress/sympathetic arousal. This may indicate that the window of time to notice changes in HRV is larger than the few minutes that comprised each of the 5 segments in loading conditions.

SCR, on the contrary, is more clearly responsive within each “loaded” segment (S2 and S4), especially for motoric loading (texting). It is likely to be a better indicator for changes in sympathetic arousal on shorter timescales.

Response to Unintended Acceleration

SCR-Amplitude showed a significant and large change on either side of the UA event, therefore is a clearly better indicator for detecting sudden, “acute stress” events. All participants who experienced a change in SCR-amplitude corresponding with that event did so within approximately 2.5 seconds. Whether this is sufficient time to offer mitigation strategies is yet to be determined, but considering that HRV as measures in the current study takes at least 100 seconds to provide a reliable indicator of change, it is clearly a better option for detecting acute stress events. Further analyses of the time window for measuring HRV may provide additional insights. For example, absolute heart rate may increase dramatically following an acute stress event. Perhaps measures of HRV over both short and longer time windows can provide additional insight into detecting acute and longer-term stressors.

II.3 Detection of An Extreme Phenomenon of Mental Workload: Cognitive Redline

The dissertation primarily focuses on the evaluation of displays under high workload situations, so it's naturally to extend our efforts to identify cognitive redline, the occurrence of which indicates mental overload (Wickens, Hollands, Parasuraman, & Banbury, 2012). As mentioned in Chapter I, how to detect cognitive redline is an important problem but haven't been satisfactorily solved. The previous studies showed that cognitive redline can be indicated by the degradation of multitask performance and the subjective rating scores that are higher than the predefined upper limits (Grier et al., 2008). However, these soluations based on performance measure and subjective self-reported measure of mental workload have their own limitations (see Appendix 1).

Physiological measure, another measure of mental workload, may be able to provide reliable indicator of cognitive redline. It was found that several physiological variables, including respiration rate, heart rate, and skin conductance level, approached their upper limits in the driving conditions that engaged high mental workload tasks (Mehler, Reimer, Coughlin, & Dusek, 2009). Event-related potential (ERP) also approached a plateau when the visual load was overwhelmed (Vogel & Machizawa, 2004). These findings suggest that some physiological patterns (plateau or asymptote) may link to the appearance of cognitive redline.

Inspired by these findings, a study as an offshoot of my research (Rodriguez Paras, 2015; Rodriguez Paras, Yang, Tippey, and Ferris, 2015) explored the use of physiological measures to indicate cognitive redline. In the study, we used Multi-Attribute Task Battery

(MATB-II), a NASA-developed multitask simulation tool, to manipulate five levels of mental workload: low, medium, high, very high, and extremely high. Cognitive redline was expected to appear among the high, very high, and extremely high workload conditions.

We found that HRV (indicated by pNN50) was higher in the low and medium workload conditions but went down and stayed reliably at a lower level when the workload increased from high to extremely high level. Meanwhile, NASA TLX went up consistently from easy to difficult and also from difficult to extremely difficult workload conditions, showing different momentum than HRV after the high workload condition. NASA TLX demonstrated the effectiveness of the manipulation of perceived workload. In addition, MATB-II didn't change the level of stress before and after the study according to short stress state question (SSSQ; Helton, 2004), which excluded the effect of stress on HRV (Matthews & Campbell, 2010). Combing the subjective (NASA TLX and SSSQ) and objective (HRV) findings, we concluded that HRV may be a reliable indicator of cognitive redline in multitask environments.

The following studies of this topic include the validation of using the asymptote patterns of HRV to indicate cognitive redline under different task environments and the test of different physiological indicators. Our long-term goal is to develop a system that can predict cognitive redline in a real-time manner and trigger effective countermeasures, such as adaptive automation, to assist operators' mental workload management in high workload situations, improving operational safety.

II.4 Conclusion

In order to measure cognitive efficiency, this chapter discussed the measurements of its two dimensions: display informativeness and imposed mental workload. The first section of this chapter explained the calculation of information transmitted as the reliable quantitative indicator of display informativeness. However, the calculation required participants' additional response (e.g., verbally report or button pressing), which might interrupt their primary task performance.

In the second section, two physiological measures (HRV and GSR) were selected to evaluate the drivers' cognitive states under different short-term loads and unintended acceleration in driving. The results showed that HRV was a sensitive indicator to short-term but prolonged loads and SCR was a more immediate indicator of an acute stress event. However, the sensitivity of these physiological measures can be affected by many factors, such as the complexity of mental workload, the predefined time windows for data analysis, and the surrounding physical environments.

In the end, this chapter discussed the methods for detecting cognitive redline - an extreme situation of mental workload - and found that the lower limit of HRV may be an effective indicator of cognitive redline, which needs further validation. Ultimately, all these efforts will contribute to affordable and high-quality evaluations of cognitive efficiency of displays under various conditions.

CHAPTER III

THE INITIAL INVESTIGATION OF COGNITIVE EFFICIENCY*

In Chapter I, one solution to data overload problem is the use of ambient displays which support the distribution of processing resources as well as minimizing overall mental workload. However, the designers of ambient displays often seem to be motivated as much or more so by aesthetic or artistic qualities than by the quality of "calmness" (e.g., Pousman & Stasko, 2006). Moreover, it is still not clear how the engagement of additional ambient displays might affect operators' mental workload.

Naturally, it can be assumed that the larger amount of cognitive resources are required for the processing of more information. But it's interesting to ask what the corresponding increase of required mental resources is, given the increased amount of information content. Since ambient display aims to present more information while imposing relatively lower mental workload, we may argue that it is more *cognitively effici*[1]*ent*, which is a critical display characteristic for addressing data overload problems.

The research in this chapter is motivated by the need to better quantify the efficiency of displays based on the methods described in Chapter II. The studies in the past have attempted to quantify the *effectiveness* of displays, for example, using signal-to-noise ratios to estimate the data available in a display (e.g., Darkow & Marshak, 1998). In

---

contrast, the study in this chapter seeks to answer **how we can measure the cognitive efficiency of basic displays in single-task conditions,** which is the second research question (Q2) of the dissertation.

III.1 The First Version of Cognitive Efficiency Metric

The initial Cognitive Efficiency (CE) metric consists of the measure of each dimension of cognitive efficiency. Inspired partly by the measure of instructional efficiency, the metric is constructed as the ratio of display informativeness (i.e., information reliably communicated by a display) and required mental resources (e.g., the amount of mental resources required to process that information). The initial CE metric is shown in Equation 5 below.

$$\mathrm{CE} = k\,\frac{display\ informativeness}{required\ mental\ resources} \tag{5}$$

In Equation 5, *display informativeness* can be evaluated based on the sensory, perceptual, and behavior response; however, they cannot provide quantitative indicator of the amount of perceived information. Thus, display informativeness in the dissertation is indicated by information transmitted (i.e., the amount of information that is present in the stimulus and reliably propagated to the observer), as defined by Information Theory (Miller, 1954; Shannon, 1948). The details of information transmitted calculation were explained in Chapter II.

In this study, we used two types of methods to estimate the *required mental resources* in process displayed information: 1) self-reported subjective workload measures (e.g., NASA-TLX ratings; Hart & Staveland, 1988) and 2) physiological measures that have been shown to correlate with mental workload (e.g., ElectroEncephaloGraphy (EEG) and skin conductance) (O'Donnell & Eggemeier, 1986).

In addition to the two dimensions, the variable k in Equation 5 represents a set of contextual factors under which the operator processes the display, such as the levels of physical energies (e.g., light, sound, etc.) in the environment, and the demand for mental resources imposed by concurrent tasks. k may be expressed as a quantitative vector of processing demands for various resources (which could be empirically derived or estimated with the aid of models). However, when multiple displays are evaluated within the same context, k can be considered to be constant and need not be further considered. In such cases, reliable comparisons can be made among CE ratios for these displays, but they are only valid within the evaluation context.

The present study represents the earlier effort of developing a robust measure of cognitive efficiency. The CE equation was used to evaluate visual and auditory displays which encoded information by modulating very basic signal dimensions (i.e., intensity and spectrum). Information transmitted by these displays over brief time intervals was calculated, and subjective (NASA-TLX ratings) and physiological (EEG and skin conductance) response were collected during the same intervals to infer the mental resources required to process the displayed information.

The results of this study are useful in several ways. First, we can see the effects of display modality and signal dimension on the information transmission between a display and a human receiver in a given context. We can also see the effects of these display characteristics on the mental workload measures, and we can compare the effects between the different workload measures. These outcomes are valuable for comparison to existing evaluations and to inform the design of displays that are intended to be highly informative and/or minimally impact processing resources, such as ambient or peripheral displays.

Future work will build on these results toward the continuing development of a robust and universal CE metric. Such a metric could be applied to any system in which human information processing is critical to safe/effective operation and where human mental resources are especially in demand. Prototype displays could then be evaluated for their predicted effectiveness in the given environment by comparing the CE ratios, and those that score the highest represent the most promising means of combating the risk of data overload.

III.2 The Initial Investigation of CE metric

III.2.1 Method

Twenty-five engineering students from Texas A&M University participated in this study. All were healthy adults with normal or corrected-to-normal visual and auditory acuity, and no known color vision deficiency.

Displays

Two visual and two auditory display prototypes were designed for this study. Within each type of media, the two displays modulated different basic signal dimensions: the intensity of the signal (denoted by the related psychological properties of light brightness and sound loudness); and the spectral qualities of the signal (light hue and sound pitch) (see Table 7 below). The displays changed state by moving between levels within the modulated dimension continuously.

Table 7: Four Displays Evaluated in the Present Study

	Visual displays	**Auditory displays**
Intensity modulation	*B*: brightness	*L*: loudness
Spectrum modulation	*H*: hue	*P*: pitch

The visual displays included clustered tri-color RGB LEDs covered by a diffusing frosted globe, and were driven by an Arduino microcontroller via MATLAB. This allowed controlling the hue and intensity of the light via RGB codes. The auditory display signals were delivered via noise-cancelling earbuds and were also generated in MATLAB by specifying the frequency, gain, and duration of each tone.

Versions of each of these displays included 2, 3, or 4 levels along the modulated dimension. In order to standardize the "psychological distance" between levels when there were more than 2, the brightness for the *B* display were set using Stevens' Power Law with an exponent of 0.45 (derived empirically via pilot tests). A calibration procedure was done for each participant so that the levels of auditory intensity in *L* were matched in terms of "subjective salience" to the levels of brightness in *B*. For the spectral displays, the levels of *H* were equidistant changes along a continuum from a pure blue hue to a pure red hue (with purplish gradations between), and the levels of *P* were equidistant pitch changes, which also assured that there would be less confusion due to harmonic resonance. Table 8 summarizes the level settings for each display.

Table 8: Levels of Each Perceptual Dimension for each Display

Display	2 levels	3 levels	4 levels
B: Brightness (index for voltage applied: 0-255)	15 59	15 95 255	15 59 138 255
L: Loudness (% maximum gain)	20 40*	20 60 100*	20 40 80 100*
	*note: these are typical values but were calibrated for each participant to match the salience of *B*		
H: Hue (RGB codes)	(0,0,20) (175,0,255)	(0,0,20) (175,0,255) (255,0,80)	(0,0,20) (175,0,255) (255,0,80) (255,0,0)

Table 8: Continued

Display	2 levels	3 levels	4 levels
P: Pitch (frequency in Hz)	329.62 415.30	261.62 369.99 523.25	261.62 329.62 415.30 523.25

Procedure

After consenting to participate, the participants were introduced to the displays and data collection methods, then completed the display calibration. They then completed 12 experimental blocks, one for each type of display at each number of levels, with the block order balanced between participants. Each block began with a 3-trial training session which could be repeated until the participant felt comfortable with the display. Then, 12, 18, or 24 experimental trials were completed for 2-, 3-, or 4-level displays, respectively.

Each experimental trial consisted of a presentation in which the display frequently changed between levels in a quasi-random fashion, with the duration at each level between 1 and 5 seconds and separated by a short (250 ms) break. The level transition/repeat probability was balanced, so the next level to be presented could not be predicted by knowing the current level. The duration of each complete trial was between 7s and 13s, with an average of 10s.

Participants were instructed to observe the display, while otherwise attempting to relax as much as possible to keep the EEG measures of attention low. For each trial, when the presentation was completed and the display went dark/silent, participants were instructed to respond as quickly but accurately as possible by clicking the button corresponding to the last (first 1/3 of trials in each block), second-last (middle 1/3), or third-last level presented (final 1/3). Participants were reminded prior to the trial which level was of interest. This N-back task component was deemed necessary in order to increase the working memory load and to avoid ceiling effects in response accuracy. Clicking the response button would display a short presentation of that level, and participants were allowed to change their response until they pressed a separate button to begin the next trial.

Participants took short rests between experimental blocks, and upon completing all 12 blocks, completed a short debriefing survey. The entire duration of the experiment was approximately two hours for each participant.

Data Analysis

By creating a stimulus-response confusion matrix from participant responses, the information transmitted (denoted as H_T) was calculated by following information theory methods (e.g., Miller, 1954; Shannon, 1948). H_T therefore represents the amount of information that is reliably communicated by the display to the human, and was used as the measure of *display informativeness* for the CE equation (5).

The *required mental resources* component of the CE equation was measured two ways, for the sake of comparison. Subjective measures of workload were gathered using the NASA-TLX survey method (Hart & Staveland, 1988), with weightings of the components calculated at the end of the study and ratings for each component collected immediately following each experimental block. Limits of the ratings scales were set such that the final TLX indices were between 0.0 and 10.0, within the same order of magnitude as the H_T measures.

Physiological measures were collected as a more objective indication of *required mental resources*, including skin conductance (data not reported here) and EEG data. For EEG, an off-the-shelf toy system (Mindflex™, developed by Mattel) was modified with the use of an Arduino microcontroller to allow collection of EEG data with minimal cost and intrusiveness to participants. For this study, “Att” (from “Attention”) is an index derived from the amount of EEG activity measured over the frontal cortex in the theta spectral power range (4-7 Hz), which, when measured over the frontal cortex has been shown to increase with increased task difficulty and higher memory load (Parasuraman & Caggiano, 2002). Att data were filtered so as to only consider the values collected while displays were being observed (not during responses or between trials), and mean Att was calculated for each experimental block. As with TLX, the Att index was also scaled to have a minimum value of 0.0 and a maximum value of 10.0.

For this study, two CE ratios were calculated that related H_T to each workload measure (H_T/TLX and H_T/Att), for each participant in each experimental block. Because large interindividual differences were expected for each individual measure, repeated measures ANOVAs were used to test for main and interaction effects for each CE ratio as well as the individual components of the ratios, with independent measures of *#levels* (2, 3, or 4), *modality* (A: auditory, or V: visual), and *dimension* (I: intensity, S: spectrum). Post-hoc t-tests with Tukey corrections were used to analyze differences among means for significant effects.

III.2.2 Results

Data for one participant were removed from analysis due to a failure that prevented saving the EEG data. The data for the remaining 24 participants were analyzed in repeated measures general linear models (formulated in Minitab).

#levels

H_T, TLX indices, and Att indices all consistently increase when the number of levels present in each display increases from 2 levels to 4 levels. Table 9 summarizes the mean values and significance.

Table 9: Main Effect of #levels on Primary Dependent Measures

	2 levels	3 levels	4 levels	Sig.
H_T (bits)	0.80	1.10	1.26	p<0.001
TLX (index)	4.10	4.98	5.83	p<0.001
Att (index)	1.79	2.78	3.55	p<0.001

Of the two CE ratios, *#levels* is only significant for H_T/Att ($F(2,253)=8.278$; $p<0.001$). Post-hoc tests show that H_T/Att decreases consistently from 2 levels to 4 levels, with the mean for 4 levels (mean=0.38 bits/index) being significantly lower than the mean for 2 levels (mean=0.52; $p<0.001$) and for 3 levels (mean=0.46; $p=0.014$). The means for 2 and 3 levels did not differ significantly. See Figure 10 for the values of each CE ratio for displays with 2, 3, and 4 levels.

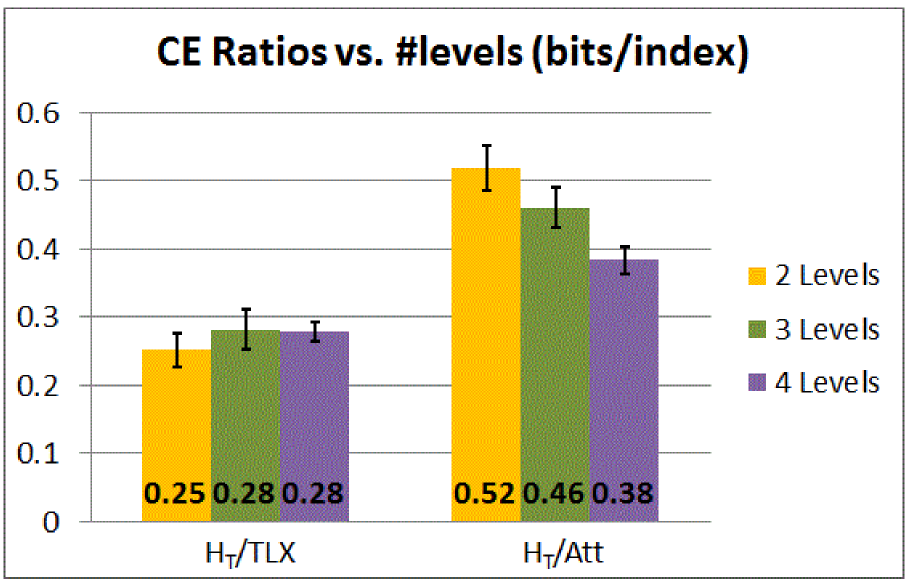


Figure 10. HT/TLX and HT/Att for displays with 2, 3, and 4 levels. Error bars represent standard error.

Modality

Comparing the auditory (A) and visual (V) displays (see Table 10) shows that visual displays transmitted more information (higher H_T), and had lower subjective workload ratings (TLX) than auditory displays. Interestingly, the Att indices showed the opposite pattern, with significantly higher values for visual displays, compared to auditory.

Table 10: Main Effect of Modality on Primary Dependent Measures

	A	V	Sig
H_T (bits)	0.94	1.17	p<0.001
TLX(index)	5.29	4.65	p<0.001
Att (index)	2.35	3.07	p<0.001

Modality has a significant effect on both CE ratios: H_T/TLX ($F(1,253)=24.51$; $p<0.001$) and H_T/Att ($F(1,253)=4.94$; $p=0.027$), however, the effects are in opposite directions. H_T/TLX for A (mean=0.21 bits/index) is lower than that for V (mean=0.33). In contrast, H_T/Att for A (mean=0.48) is higher than that for V (mean=0.42).

A significant interaction effect was found between *#levels* and *modality* for H_T/Att ($F(2,253)=5.80$; $p=0.003$). Figure 11 shows that for displays with 2 or 3 levels, auditory displays outperform visual displays. Post-hoc tests show a marginally significant difference at 2 levels ($p=0.078$) and a significant difference at 3 levels ($p=0.013$). For displays with 4 levels, although there is no significant difference, we can see at least a trend that shows visual displays outperforming auditory displays, which is in line with the findings for H_T/TLX ratio across *#levels*.

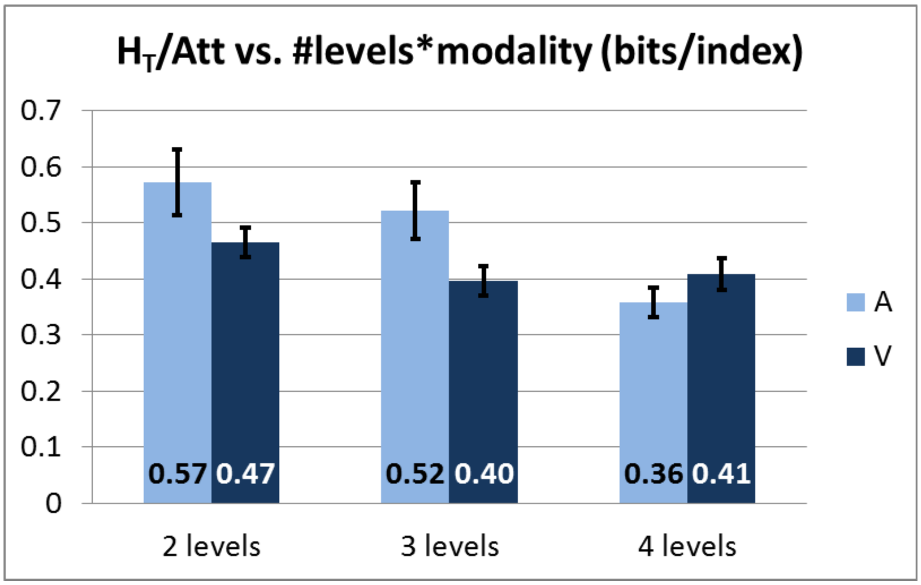


Figure 11. Interaction effect between #levels and modality on HT/Att. Error bars represents standard error.

Dimension

The display dimension which was modulated to encode the data, intensity (I) vs. spectrum (S), was a significant factor for H_T and TLX, but not for Att indices. Spectrum encoding resulted in more information transmitted and lower TLX indices than did intensity encoding. The mean values and significances are listed in Table 11.

Table 11: Main Effect of Dimension on Primary Dependent Measures

	I	S	Sig
H_T (bits)	0.91	1.20	p<0.001
TLX (index)	5.10	4.84	p=0.033
Att (index)	2.72	2.69	Not Sig

Dimension is significant for both CE ratios: H_T/TLX ($F(1,253)=8.66$; $p=0.004$) and H_T/Att ($F(1,253)=15.41$; $p<0.001$). In each case, S encoding (H_T/TLX mean=0.30 bits/index; H_T/Att mean=0.51) results in more efficiency than I encoding (H_T/TLX mean=0.23; H_T/Att mean=0.23).

For H_T/TLX, an interaction effect was found for *#levels* and *dimension* ($F(2,253)=4.19$; $p=0.016$) (see Figure 12). Post-hoc tests showed that for displays with 2 and 3 levels, no significant difference is found between I and S encoding methods. However, for 4 levels, S encoding (mean=0.37 bits/index) significantly outperforms I encoding (mean=0.19; $p<0.001$).

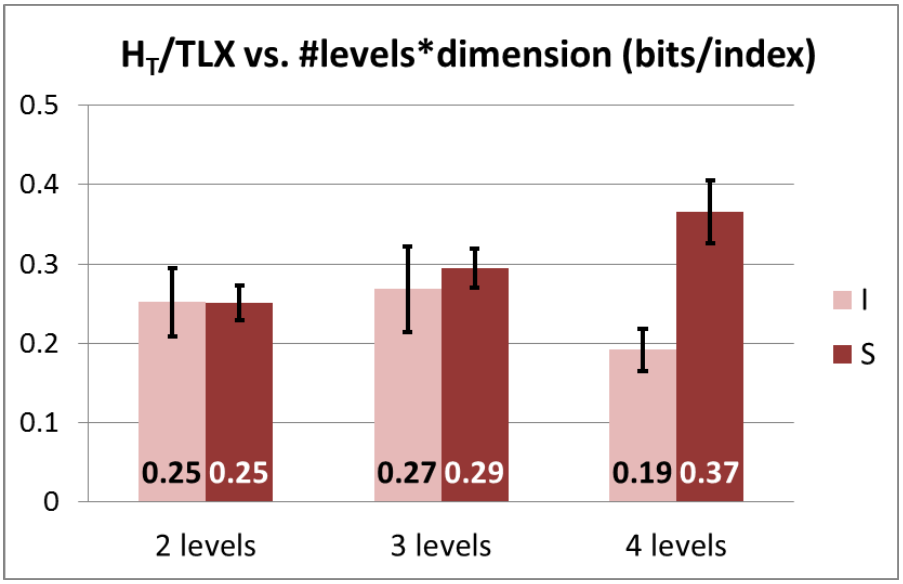


Figure 12. Interaction effect between #levels and dimension on HT/TLX. Error bars represents standard error.

III.2.3 Discussion

This chapter introduces a new way - Cognitive Efficiency (CE) metric - to evaluate display design. The metric is formulated as the ratio of display informativeness and required mental resources, affected by the contextual factor *k*. Since required mental resources can be measured subjectively or objectively, and each method has its strengths and limitations (e.g., O'Donnell & Eggemeier, 1986; Wickens, Hollands, Parasuraman, & Banbury, 2012), we measured it in both ways for the sake of comparison. The study in this chapter evaluated the cognitive efficiency of very basic visual and auditory displays which were designed to be similar to existing ambient displays.

The results showed that, as expected, when information content (*#levels*) increased, so did the measures of mental workload. Interestingly, however, the H_T/TLX ratio did not change significantly due to *#levels*. This might suggest that the H_T/TLX ratio is less sensitive to changes in information content, since it remained relatively constant within each type of display. The fact that the H_T/Att ratio consistently decreased with increasing information content may mean that this objective measure is more sensitive to increased workload, and may also reflect the functional relationship between EEG measures and workload.

It was interesting to notice that cognitive efficiency was different between the visual and auditory displays. Mirroring the general recommendation that displays with higher information content should be displayed visually (e.g., Wickens, Hollands, Parasuraman, & Banbury, 2012), the H_T/TLX (subject) ratio reflected visual displays had greater cognitive efficiencies than auditory displays. The H_T/Att (objective) ratio, in contrast, suggested auditory display was more cognitively efficient than visual display for lower information content, and visual more efficient for higher content. The findings may associate with the difference between perceived and actual loadings on the visual and auditory systems. While participants may have felt the auditory displays were harder to interpret (greater subjective workload and lower H_T/TLX ratios), the neural activities in cerebral cortex associated with listening to simple auditory signals are apparently less (denoted by lower ATT) than that of watching simple visual displays. In any case, these findings suggest that visual displays are more efficient *at least* with higher information content.

Finally, we found a significant effect of display *dimension* on both CE ratios. It showed that the display communication is more efficient by encoding information in the spectral (i.e., hue or pitch) than in the intensity (i.e., loudness or brightness). Additionally, dimensional difference in the H_T/TLX measure was most pronounced with higher information content. This difference may be partially due to human's lower capability of identifying the absolute levels of signal intensity (e.g., Hsia, 1971). Because the experimental paradigm required an absolute judgment in its response, it may have been more difficult for participants to distinguish between "brightness level 1" and "brightness level 2", since the levels of brightness are likely to be processed in a more relative fashion (e.g., "brighter", "less bright"). In contrast, distinguishing red from reddish-purple or bluish-purple is likely easier.

Taken together, the results suggest that for the design of simple displays that follow the ambient display model, the display should modulate spectral dimensions, rather than intensity. Visual displays are likely better for this purpose; however, auditory displays may be better for lower information content. As a general disclaimer, the results are only valid and applicable within contexts that resemble the current study, in regard to environmental stimuli and the nature of task demands.

III.3 Conclusion

Chapter III documented the initial efforts in the development of a metric to quantify the "cognitive efficiency" of (ambient) displays that present each piece of information at one moment. The results demonstrated that cognitive efficiency is different among displays that engaged different perceptual modalities or dimensions. Specifically, the cognitive efficiency of the visual displays was higher than that of the auditory displays in higher information content. Moreover, the displays were more cognitively efficient when they encoded information into the spectrum dimensions than the intensity dimensions. However, the study in this chapter only focused on basic displays under the lower task loads; the next chapter will examine the measure of cognitive efficiency in the high-load multitask scenarios and its relationship with multitask performance.

CHAPTER IV

THE RELATIONSHIP BETWEEN COGNITIVE EFFICIENCY AND MULTITASK PERFORMANCE

Given the initial version of Cognitive Efficiency metric in Chapter III, this chapter proposes its modified version for quantifying the cognitive efficiency (CE) of displays in human-machine systems by taking account into human, display, and contextual factors. Moreover, this chapter examines correlations between the CE metric and multitasking performance in a driving simulation. The modified CE metric uses existing theory and methods to quantify both *display informativeness* and *required mental resources*. A divided-attention task set involved processing different visual displays to inform route selection while concurrently avoiding obstacles in a simulated driving study. Measures of multitasking performance as well as *informativeness* and *resources required* were collected while participants processed each display. Correlation analyses were used to identify the relationships between five constructed CE metrics and performance that differed between High- and Low-performing groups, potentially attributable to differences in imposed workloads and/or cognitive "redlines".

IV.1 The Modified Version of Cognitive Efficiency

In Chapter III, we proposed two basic dimensions that determine the cognitive efficiency of displays: 1) *display informativeness* (amount of displayed information accurately processed by the human); and 2) the types and amounts of *required mental resources* consumed in the specific act of processing this information. The cognitive efficiency of displays values communicating more information per "unit" of required operator mental resources (Yang, Shukla, & Ferris, 2012).In Chapter III, the initial investigation found that simple displays (similar to ambient displays) showed different levels of cognitive efficiency, which conformed with expectations according to information processing literature (Yang et al., 2012).

In addition to the basic displays evaluated in Chapter III, Chapter I listed several other display design paradigms that also offer insight into the improvement of cognitive efficiency. For example, Ecological Interface Design (EID) reduces mental efforts by supporting minimally-demanding skill-based and rule-based behaviors and providing structured information to ease knowledge-based reasoning (Vicente, 2002). Multimodal displays can distribute processing requirements across multiple perceptual channels to avoid overloading them (Sarter, 2006; Wickens, 2002). Redundant multisensory displays also reduce the risk of overload by supporting flexibility in engaged channels (Wickens, Prinet, Hutchins, Sarter, & Sebok, 2011) but may hinder performance due to increased signal complexity (Ardoin & Ferris, 2016). "Preattentive reference" displays minimize the required mental effort and risk of concurrent task interference when processing them (Woods, 1995).

In addition to display format, human and contextual factors also contribute to cognitive efficiency. Human factors can include operator information processing abilities, preferences, and mental/emotional state. Some factors are stable over time, such as whether one is a "visual" or "verbal" learner (e.g., Mayer & Massa, 2003), while emotional state and stress levels can also temporarily affect mental storage and processing capacity (Eysenck & Calvo, 1992; Eysenck, Derakshan, Santos, & Calvo, 2007). Contextual factors defined by the environment or task set can also impact information processing, and can interact with human factors to affect efficiencies in complex ways.

In this chapter, we modified our previous version of Cognitive Efficiency metric and propose a new version for quantifying CE of displays in human-machine systems (Equation 6). The metric is also formulated as a ratio of *display informativeness* to *required mental resources*, but taking into account display, human, and contextual factors in a given multitasking context. The components of Equation 6 can flexibly integrate several existing theories and tools described in Chapter II.

$$\mathrm{CE}(h_i, d_i, k_i) = \frac{display\ informativeness\ (h_i, d_i, k_i)}{required\ mental\ resources(h_i, d_i, k_i)} \qquad (6)$$

The new version of CE metric illustrates how the CE metric and its components are functions that depend on characteristics of the human (*h*), display (*d*), and task/environmental context (*k*). These characteristics can be expressed quantitatively, for example, *k* could be represented by a vector of loads imposed by other concurrent tasks on various dimensions of cognitive resources, such as individual sensory channels or working memory functions (e.g., Wickens, 2002). The variables can also qualitatively describe the contexts under which CE measurements are conducted. Investigations to date have emphasized consistent experimental contexts so that k can be handled as a constant. Cognitive efficiency in context k_i can then be compared between displays (with the same human, h_i) or between humans (with the same display, d_i).

Given the modified version of CE metric, Chapter IV aims to determine how different constructed CE metrics relate to performance in a high-demand multitask set. Operators' multitask performance in human-machine systems is expected to improve when task-related displays support higher cognitive efficiency. This expectation follows from two assumptions: all other factors aside, 1) display-related task performance is the same or better with greater *display informativeness*; and 2) secondary task performance is the same or better with fewer *required mental resources* for display-related task processing, since more resources are "left over" for secondary tasks. The primary question in this chapter then becomes which of the many ways to measure *display informativeness* and *required mental resources* should be used to construct the CE metric to most accurately predict multitasking performance.

This chapter investigated the relationship between cognitive efficiency metrics and multitask performance in a simulated driving environment. Driving scenarios required performing two concurrent tasks that competed for shared mental resources to varied extents. Multitask performance was calculated by combining scores for each task. The measures of *display informativeness* and *required mental resources* were collected and used in CE metric calculations. Correlation analyses then sought to find relationships between CE constructs and multitask performance.

The long-term goal of this research is to develop a generalized technique for reliable CE measurement that is sensitive to a broad range of display, human, and contextual factors. A secondary goal is to achieve this with assessment methods that are cost-effective and that minimally interfere with natural task performance. Validated CE measurement techniques may then be used to conduct heuristic assessments under representative operating contexts and to predict the effectiveness of displays in supporting multitask performance. Additionally, this research provides basic scientific insight into multitask information processing, and design to reduce the risk of data overload in complex task environments.

IV.2 Investigating the Relationship Between Cognitive Efficiency and Multitask Performance

IV.2.1 Method

Participants

Twenty-four students and employees of Texas A&M University (18 males and 6 females, mean age 25) participated in this study. All participants had normal or corrected-to-normal visual acuity with no color vision deficiencies, and possessed a valid driver's license for at least one year, with an average of 5.5 years of prior driving experience.

Procedure

After signing the consent form and affixing physiological sensors, participants were familiarized with the driving simulator. They then demonstrated a basic proficiency in experimental tasks by completing a 15-minute training scenario with samples from each display type. Next, the participants completed four experimental conditions in a counterbalanced order. Each condition included twelve trials of tasks. Physiological data were collected continuously throughout the study. NASA TLX questionnaires (Hart & Staveland, 1988) were completed after each condition. The study lasted approximately 2 hours and each participant was compensated 20 dollars.

Driving Scenarios

Four driving simulation scenarios were created in the STISIM Drive® driving simulator, controlled with a Logitech G27 racing wheel and integrated throttle and brake pedals. Figure 13 illustrates a top-down schematic of each driving scene and an embedded example of the driver view. Each scenario consisted of a 40-foot wide one-way road with alternating Segment 1 and Segment 2. Segment 1 allowed free lateral movement over the full roadway width while Segment 2 required participants to enter one of four lanes delineated by longitudinal pylons that prevented participants from changing lanes after entering the segment (see Figure 13). A "buffer zone" between the two segments allowed for participants to navigate to their chosen lane before entering the lane segment. Participants were instructed to drive at the constant governed speed of 35 MPH, and to focus control efforts on the lateral position of the vehicle. Their goal was to minimize the number of collisions with roadway brush while remaining on the roadway. Driving off-road registered one collision for every 10 feet off-road.

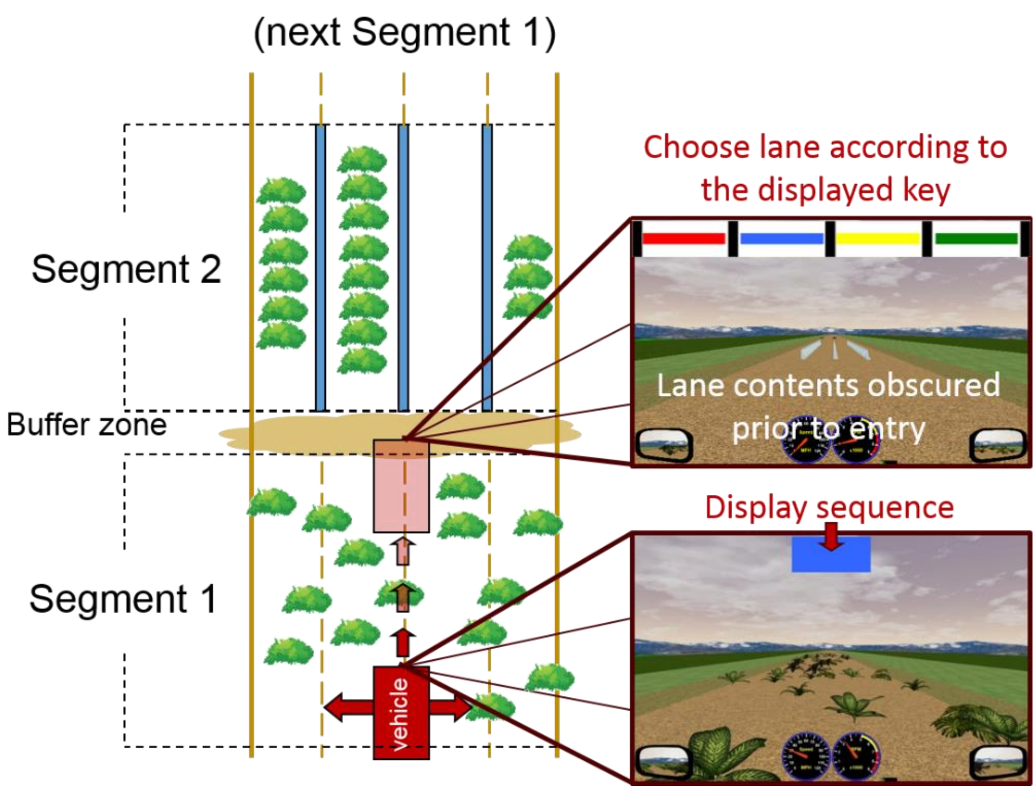


Figure 13. Schematic of the driving scenario and example driver views while in the Segment 1 and Segment 2.

Tasks

There are two types of tasks: 1) the *obstacle-avoidance task* (OAT) and 2) the *lane-selection task* (LST). The OAT was designed to continuously load the participant's visual channel and spatial working memory while navigating Segment 1. It involved observing the pseudo-randomized distribution of plants in this segment (visible up to 400 feet away; see Figure 13) and mentally planning and executing routes that minimized the number of collisions. Crushing sounds were played as feedback for each registered collision.

The LST extended from Segment 1 to Segment 2. It required participants to monitor visual stimuli on a heads-up display while navigating Segment 1, and when entering Segment 2, choose a lane that minimized brush collisions in that segment. The heads-up displays conveyed pseudo-randomized sequences of stimuli, each presented for 2 seconds with 1 second between presentations. Each stimulus represented one of four equiprobable states and were encoded according to display type. Three different experimental displays, and additionally a baseline display were employed in the four separate display conditions (see Figure 14).

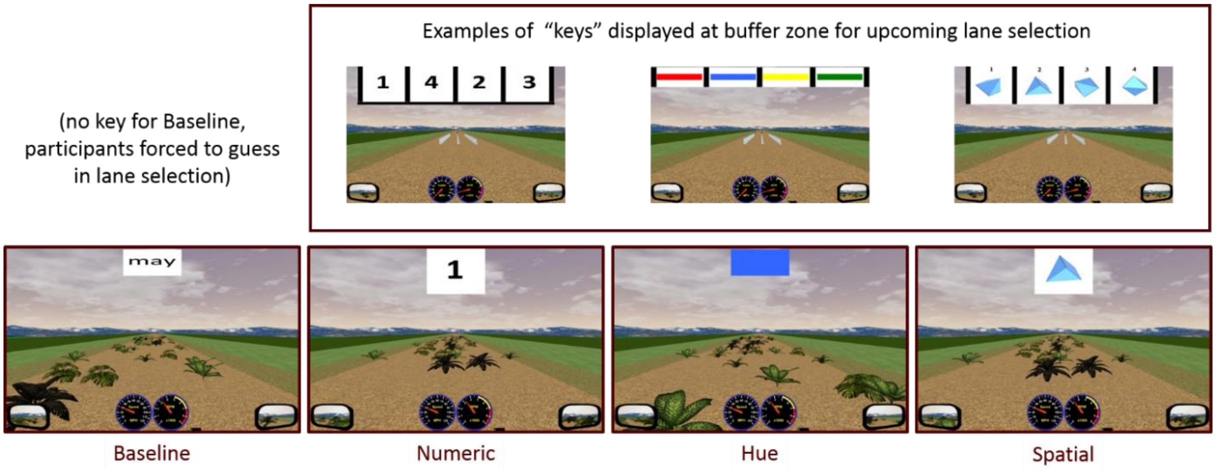


Figure 14. Four types of head-up visual displays and their respective keys: Baseline (single common words), Numeric (Arabic numerals 1 - 4), Hue (rectangles colored red, blue, yellow, or green), and Spatial (a 3-D pyramid in four distinct spatial orientations).

In the Numeric, Hue, and Spatial experimental display conditions, participants were told to remember the last three presentations of the display sequence (which included between 4 and 10 stimuli). Limited visibility prevented participants from predicting the end of the sequence, thus they were required to maintain the three latest presentations as in a running memory task. When reaching the buffer zone, the simulation would pause eight seconds so participants could verbally report, in order, the last three presentations they observed (or their best guess). This report was used to calculate the measures of display informativeness in Equation 6. Meanwhile, a "key" was presented to pseudo-randomly labeled the four upcoming lanes with associated stimuli (see Figure 15).

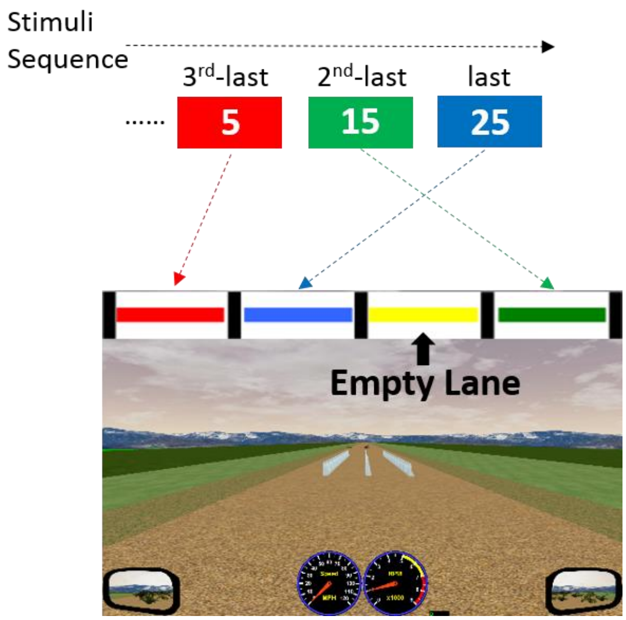


Figure 15. LST key and expected number of brush collisions in each lane.

After the eight-second pause, the simulation would resume and the participants needed to maneuver their vehicle into the chosen lane in Segment 2. The choice of lane should be motivated by minimizing brush collisions in that lane, and is informed by the last three presentations of the head-up display, each of which communicate lanes with increasing density of brush. Figure 15 provides an example in which the final three hue presentations (in order) were red-green-blue. The blue presentation was the most recent, so is associated with the lane that includes the most brush (expected 25 collisions). In this way, the most accessible display content (with regard to memory) provides the most useful information about which lanes to avoid. Choosing the lane tagged by green (last second presentation) would result in 15 collisions and the lane tagged by red (last third presentation) would result in 5 collisions. Processing all three presentations supports the identification of the yellow lane as the lane with the least brush density (0 collisions) and thus the best lane choice for overall performance.

The information processing for both tasks overlapped in Segment 1, creating a de facto concurrent task set. Participants were told to adopt any strategy that would result in the fewest brush collisions across both tasks. While balanced strategies were encouraged, extensive pilot testing was used to distribute brush in each of the "lane" and "obstacle" segments so that extreme strategies (i.e., focusing 100% on one task and 0% on the other) would result in roughly equivalent numbers of collisions.

The Baseline display condition imposed similar response-related workload by requiring verbal report of the last word from each display sequence, but Baseline stimuli carried no information regarding the LST, forcing participants to choose lanes at random.

An index of overall performance in both tasks (PI) was calculated according to Equations 7, 8, and 9. PI values ranged from 0 (worst) to 1 (best multitask performance).

$$PI_j = \left(PI_{LST\ in\ j} + PI_{OAT\ in\ j}\right) \times \frac{1}{2} \tag{7}$$

$$PI_{LST\ in\ j} = \frac{135 - NC_{LST\ in\ j}}{135 - 0} \tag{8}$$

$$PI_{OAT\ in\ j} = \frac{325 - NC_{OAT\ in\ j}}{325 - NC_{OAT\ in\ Baseline}} \tag{9}$$

$NC_{LST\ in\ j}$: number of collisions in the LST in condition j

$NC_{OAT\ in\ j}$: number of collisioins in the OAT in condition j

$NC_{OAT\ in\ Baseline}$: number of collisioins in OAT in the Baseline condition

j: experimental display condition (Numeric, Hue, or Spatial)

Note. 135 is the expected number of collisions in the LST when displays are completely ignored and participants choose lanes at random, and 325 is the expected number of collisions in the OAT when brush locations in Segment 1 are completely ignored (e.g., with random steering control inputs). 0 and $NC_{OAT\ in\ Baseline}$ represent the best possible performance scores in the LST and OAT, respectively, achieved by maximally allocating mental resources to the respective task.

CE Component Measures

In this study, *display informativeness* was measured in two ways: 1) percent accuracy of the verbal report (ACC) and 2) bits of information transmitted (INF) evidenced in the verbal report. The information transmitted was calculated based on the formula of entropy in Information Theory (Miller, 1953; Shannon & Weaver, 1949).

Required mental resources were represented as scaled Z-scores of mental workload indices derived from NASA TLX subjective ratings and physiological measures of skin conductance level (SCL), frontal lobe Electroencephalography (EEG), and heart rate variability (HRV) (see Table 12). The physiological data were collected precisely when participants were actively processing the displayed information. A physiological baseline collected during a three-minute relaxation period prior to beginning experimental conditions was used to scale physiological data.

Table 12: List of Physiological Measures Used in Required Mental Resources Quantifications

Physiological Measure (Device)	Sensor Location	Measurement Index	Index Meaning	Relationship to Mental Workload
SCL (Iom® Wild Divine biofeedback sensor system)	Tips of index and ring fingers of the non-dominant hand	The average skin conductance level over display-processing interval	Tonic phonemena of electrodermal activity (Boucsein, 2012; Mehler, Reimer, & Coughlin, 2012)	Higher SCL associates with higher mental workload (e.g., Boucsein, 2012; Mehler, Reimer, & Coughlin, 2012)
HRV (Zephyr Bioharness 3)	Under participants' clothing around the torso	pNN20 and pNN50	The percentage of successive N-N intervals that differ by more than 20ms and by more than 50ms	Lower pNN20 or pNN50 associates with higher workload (e.g., Cinaz, Arnrich, La Marca, & Tröster, 2013; Mehler, Reimer, & Wang, 2011).

Table 12: Continued

Physiological Measure (Device)	Sensor Location	Measurement Index	Index Meaning	Relationship to Mental Workload
EEG (NeuroSky® MindWave hardware)	ingle electrode on the Fp1 position on ventrolateral prefrontal cortex (VLPFC) according to 20-10 international system	The average desynchronization percentage (ERD%) of lower alpha band of the EEG	The extent of depression from the average of EEG amplitude in the physiological baseline	Larger ERD% of lower alpha band indicates more overall mental demand (e.g., Klimesch, 1999)

The mental workload imposed by LST and OAT was calculated in Equation 10.

$$MW_{LST\ in\ j} = MW_{Total\ in\ j} - MW_{\ Total\ in\ Baseline} \quad (10)$$

j: experimental display condition (Numeric, Hue, or Spatial)

Note. MW represents mental workload.

Correlating CE Measures with Multitask Performance

Quantitative measures of display informativeness and required mental resources were compiled for each participant. Pairs of measures were then combined to construct several CE metrics according to Equation 6. Finally, correlation analyses determined the relationship between these CE metrics and multitask performance.

IV.2.2 Results

Twenty-four participants were divided into High and Low groups according to PI_{LST} in the Spatial condition by applying k-means clustering with internal validation measures (connectivity, Sillhouette Width, and Dunn Index) (see Table 13).

Table 13: Clustering of High and Low Groups

Groups	PI_{LST} in Spatial Condition	Subject Number
High	0.56 ~ 1.00	n=17
Low	0.00 ~ 0.33	n=7

Note. PI_{LST} represents performance index of the lane-selection task (see Equation 8).

Repeated-measures two-way ANOVAs were used to determine main effects of *Display*, *Group*, and their interaction effect, with significance levels of $\alpha=0.05$. *Display* is a variable that specifically included the 3 experimental display conditions: Numeric, Hue, and Spatial. Tukey post hoc tests were used to find differences among means. Finally, Pearson product-moment analyses determined correlation coefficients between CE metrics and multitask performance indices. All analyses were completed in R 3.2.3.

Performance Indices

A significant main effect of *Display* was found on PI ($F(2,66)=7.60$, $p<0.001$, $\eta^2=0.11$. The post hoc test showed that the Spatial condition (PI=0.75) involved significantly lower PIs than the Numeric (PI=0.89, $p<0.001$) and Hue (PI=0.85, $p=0.030$) conditions. The Numeric and Hue conditions didn't differ from each other. PI of the High group (mean PI=0.89) was significantly higher than the Low group (PI=0.67; $F(1,66)=47.13$, $p<0.001$, $\eta^2=0.34$).

The interaction effect between *Display* and *Group* was significant as well ($F(2,66)=5.03$, $p=0.009$, $\eta^2=0.07$). The *Display* effect was insignificant in the High group (PI: Numeric=0.93, Hue= 0.89, Spatial=0.86). However, in the Low group, PI in the Spatial (PI=0.50) condition was lower than in both Numeric (PI =0.79, $p<0.001$) and Hue (PI=0.74, $p=0.003$) conditions. The *Group* effect was significant in the Spatial condition (PI: High=0.86, Low=0.50, $p<0.001$), but insignificant in the Numeric (High=0.89, Low=0.74) and Hue (High=0.93, Low=0.79) conditions (see Figure 16).

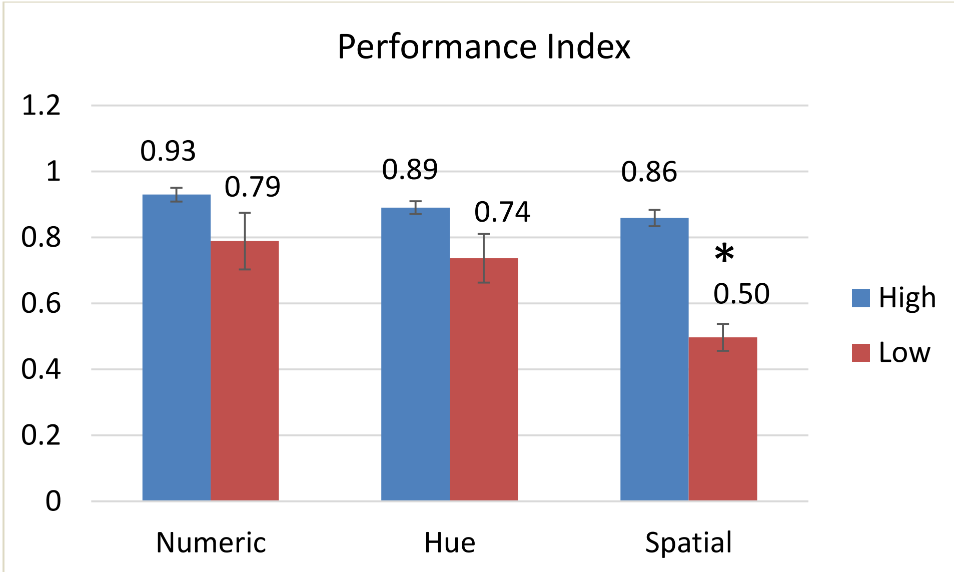


Figure 16. Performance indices for each group in each experimental display condition. Error bars represent standard error. * indicates significant difference.

Display Informativeness Measures

INF (F(2,66)=4.70, p=0.012, R^2=0.09 and ACC (F(2,46)=7.55, p=0.014, η^2=0.08) were significantly affected by *Display* (see Figure 17). Post hoc tests showed that the Numeric display (INF=1.68 bits) transmitted significantly more information than the Spatial (INF=1.38 bits, p<0.009) but neither the Numeric nor Spatial displays differed from the Hue display (INF=1.52 bits). Similarly, report accuracy was also significantly higher in the Numeric (ACC=88.0%) compared to the Spatial (ACC=75.1%, p=0.010) conditions, but neither the Numeric nor Spatial displays differed from the Hue (ACC=82.8%).

The High group (INF=1.66, ACC=88.2%) showed higher INF and ACC than the Low group (INF=1.20, $F(1,66)=27.85$, $p<0.001$, $\eta^2=0.26$; ACC=67.0%, $F(1,66)=30.31$, $p<0.001$, $\eta^2=0.27$). The interaction effect between *Display* and *Group* was not significant on either INF or ACC.

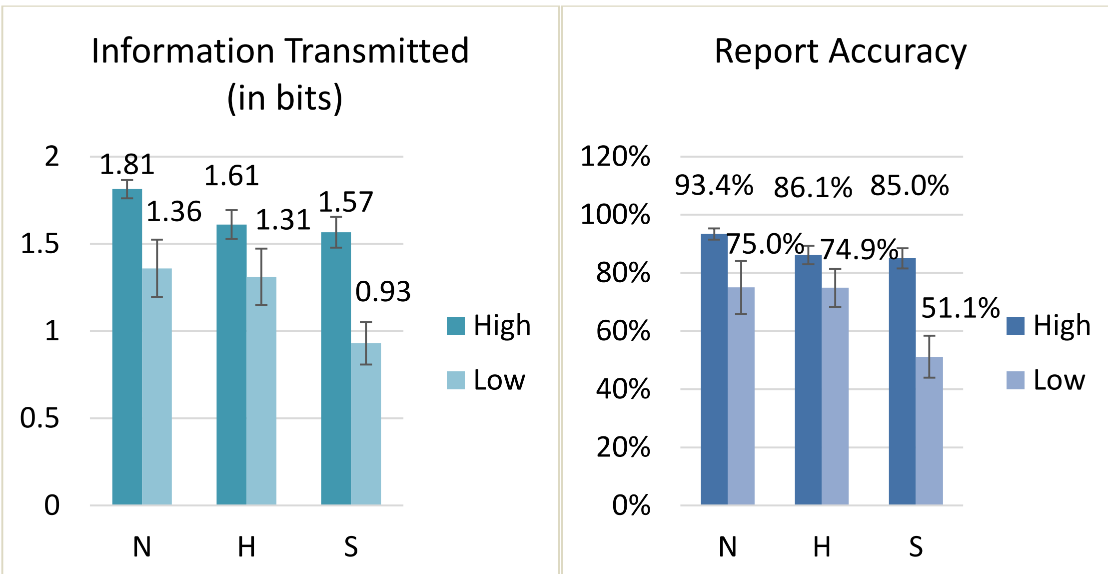


Figure 17. Information transmitted (INF; left) and report accuracy (ACC; right) for each performance group in each experimental display condition. Error bars represent standard error. N, H, and S represent Numeric, Hue, and Spatial conditions, respectively.

Measures of Required Mental Resources

Table 14 provides a summary of each mental workload measure across experimental display conditions. Among them, only NASA TLX ($F(3,88)=18.15$,

p<0.001, η^2=0.37) was significantly affected by the experimental conditions. NASA TLX of Baseline condition (TLX=78) was significantly lower than that in all *Display* conditions (Numeric=130, p =0.002; Hue=133, p =0.001; Spatial=181, p <0.001).The Spatial condition showed higher NASA TLX ratings than the Numeric (p<0.001) and Hue (p<0.001) conditions. NASA TLX was not significantly different between the Numeric and Hue conditions. In addition, the *Group* effect on NASA TLX was not significant.

Table 14: Measures of Mental Workload among All Experimental Conditions

	Baseline	Numeric	Hue	Spatial	Sig.
EEG(ERD%)	24.9%	37.8%	37.8%	32.2%	no
HRV(pNN20)	63.6%	64.1%	63.2%	63.1%	no
HRV(pNN50)	30.4%	29.0%	30.9%	28.6%	no
SCL	3.62	3.88	3.73	3.87	no
NASA TLX	78	130	133	181	p<0.001

HRV (pNN50=26.8%, pNN20=60.5%) of the High group was significantly lower than that of the Low group (pNN50=36.8%, F(1,88)=7.20, p=0.009, η^2=0.08; pNN20=70.9%, F(1,88)=7.19, p=0.009, η^2=0.08). SCL (SCL=4.35) of the High groups was significantly higher than that of the Low group (SCL=2.38, $F(1,88)=15.10$, $p<0.011$, $\eta^2=0.07$).

Workload measures were partitioned according to Equation 10 to determine relative workload contributions from LST (see Figure 18). Of these partitions, only TLX_{LST} ($F(2,66)=9.37$, $p<0.001$, $\eta^2=0.21$) was significantly affected by *Display*. The post hoc tests showed TLX_{LST} in the Spatial condition ($TLX_{LST}=101.5$) was significantly higher than the Numeric and Hue conditions (Numeric: $TLX_{LST}=50.7$, $p<0.001$; Hue: $TLX_{LST}=53.2$, $p=0.001$), but it did not differ between Numeric and Hue conditions.

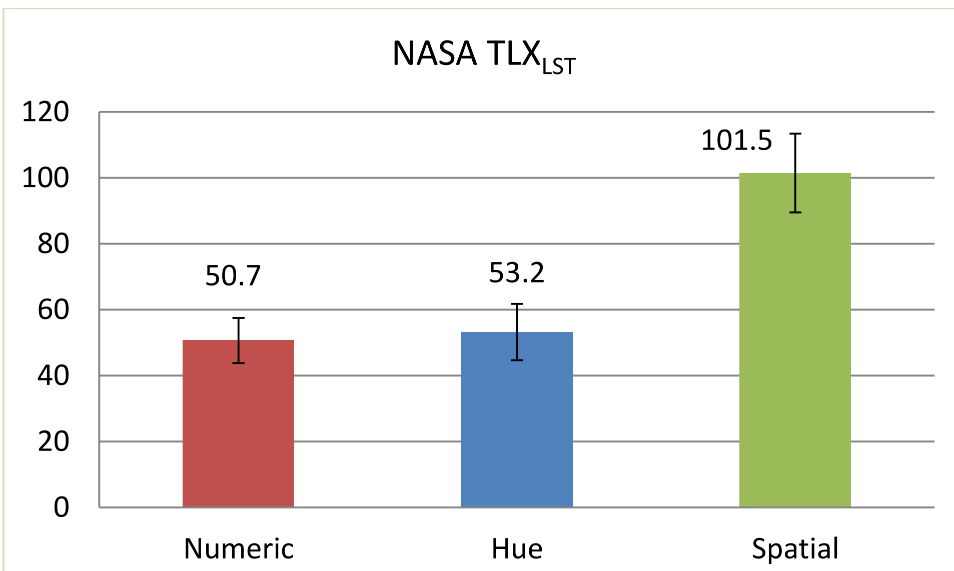


Figure 18. NASA TLX workload indices in the lane-selection task (TLXLST) for all participants in each display condition. Error bars represent standard error.

Cognitive Efficiency Metrics

Following Equation 6, five CE metrics were constructed using different combinations of measures for *display informativeness* and *required mental resources.* INF and five mental workload indices (EEG_{LST}, SCL_{LST}, HRV_{LST}(pNN50), HRV_{LST}(pNN20), and TLX_{LST}) were normalized and scaled into the same positive range. Only INF was chosen to indicate display informativeness because ACC and INF were highly correlated (r=0.97) and INF is a better indicator of the quantity of information gained from the display. As the task associated with processing the head-up displays, only the LST partitions of workload indices were considered in CE calculations.

The INF/TLX construct was the only one that was significantly affected by *Display* (see Table 15). For INF/TLX, the Spatial display scored lower than the Numeric and Hue displays. The scores of all CE metrics were significantly higher in High group than Low group (see Table 15).

Table 15: Display Effect and Group Effect on Cognitive Efficiency (CE) Metrics

	Display Effect				Group Effect		
CE Metrics	Numeric	Hue	Spatial	Sig.	High	Low	Sig.
INF/TLX	1.20	1.10	0.90	*p*<0.001 (S<Nu***, H*)	1.12	0.93	*p*=0.003
INF/EEG	1.35	1.27	1.19	no	1.34	1.27	*p*=0.005
INF/pNN50	0.97	0.93	0.83	no	0.99	0.70	*p*<0.001
INF/pNN20	1.12	1.00	0.96	no	1.13	0.78	*p*<0.001
INF/SCL	1.13	1.07	0.96	no	1.14	0.84	*p*<0.001

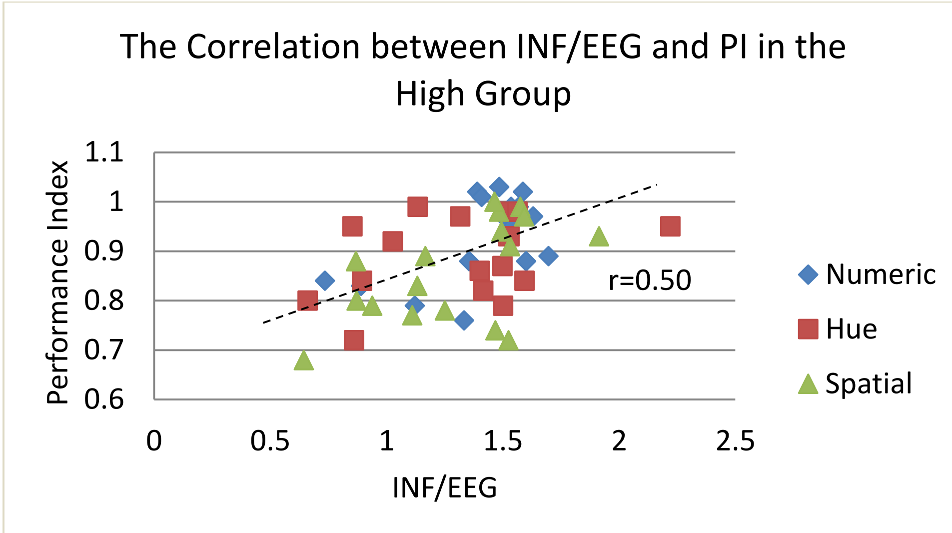
Note. Nu, H, and S represent Numeric, Hue, and Spatial display conditions, respectively. *p<.05. ** p<.01. *** p<.001.

Table 16 lists Pearson correlation coefficients between each CE metric and multitask performance indices. Most CE metrics were significantly correlative with PI. In many cases, the correlations for the Low group were significantly higher than those for the High group. Figure 19 illustrates the most highly-correlating CE metrics for the High (INF/EEG: r=0.50) and Low (INF/TLX: r=0.89) groups vs. PI in each experimental display condition.

Table 16: Pearson Correlation Coefficients between CE Metrics and Multitask Performance Indices

		All Group	High Group	Low Group	Group effect
CE	INF/TLX	0.69***	0.41**	0.89***	High<Low**
	INF/EEG	0.62***	0.50***	0.74***	no
	INF/(1-pNN50)	0.51***	0.17	0.66***	High<Low *
	INF/(1-pNN20)	0.52***	0.27	0.63**	no
	INF/SCL	0.60***	0.38**	0.80***	High<Low *
DI	INF	0.72***	0.44**	0.78***	(N/A)
	ACC	0.76***	0.42**	0.81***	(N/A)
MW in LST	TLX	-0.29	-0.11	-0.55**	(N/A)
	EEG	-0.08	-0.33	-0.18	(N/A)
	1-pNN50	-0.06	-0.11	-0.10	(N/A)
	1-pNN20	-0.08	-0.03	-0.21	(N/A)
	SCL	-0.11	-0.22	-0.04	(N/A)

Note. The workload indices in Table 16 represent the mental workload associated only with the lane-selection task (LST). DI represents quantifications of display informativeness and MW represents mental workload. ‘1-pNN50’ and ‘1- pNN20’ correlate positively with mental workload. The underlined values indicate particularly high correlation coefficients between CE and PI in each column. * p<.05. ** p<.01. ***p<0.001.



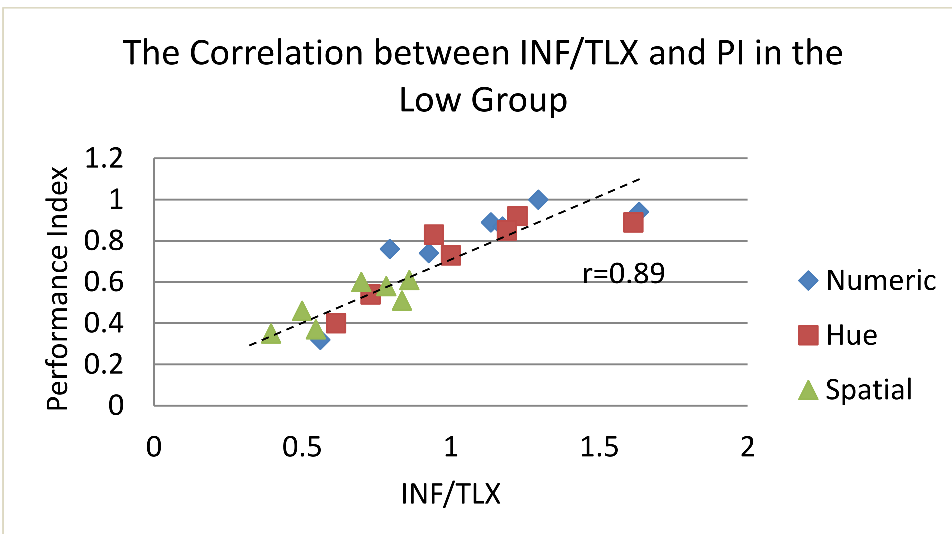


Figure 19. Relationships between CE metrics and multitask performance indices for the highest-correlating metrics for High- (top) and Low (bottom) groups.

Importantly, the CE metrics in Low group in Table 16 generally show higher correlations with PI than do the individual (display informativeness or required mental resources) dimensional measures. In the case of INF/TLX, it (r=0.89) correlated more strongly than INF (r=0.78) and TLX (r=0.55, p<0.001).

IV.2.3 Discussion

In this chapter, we define cognitive efficiency (CE) as a ratio of *display informativeness* to the amount of operator *mental resources required* in processing the displayed information. A robust CE measurement technique can be used to assess the efficiency of human-display systems in multitasking environments.

Given the complex nature of the concepts represented by the *informativeness* and *required resources* dimensions, it is challenging to find ways to quantify each dimension that are reasonably simple yet sufficiently reflect these complexities. Existing theory and measurement techniques can be used to crudely quantify each dimension, but the appropriateness and sensitivities of existing methods depend on characteristics of the human, display, and context. Therefore, a generalized construct must allow flexibility in choice of methods. This study sought to determine how closely various CE constructs and metric calculations correlated with multitask performance in a specific task context: gathering coded navigational information while driving a vehicle. These efforts revealed that CE metrics represent promising means of predicting multitask performance in similar task contexts.

Multitask performance, indicated by performance indices (PI), was improved when operators gained more information from experimental displays and used it to make better lane-choice decisions in the LST. PI also increased when displays imposed smaller demand on visual-spatial resources, thus leaving more available for route-planning and navigation in a concurrent OST. Since LST and OST contributed roughly equivalently to performance indices, PI was relatively unaffected by variations in task management strategies.

CE Dimensional Measures

In the current study, *display informativeness* was quantified in two ways: information transmitted (Miller, 1953; Shannon & Weaver, 1949) and report accuracy, which were highly correlated with each other so that they provided roughly equivalently-predictive results. Although we primarily focused on information transmitted in the paper, the relative ease of the accuracy measure may make it preferred, at least for simple displays with low overall information content. However, both measurements of display informativeness require verbal report so that they may interrupt the driving performance. Therefore, there are needs of exploring unintrusive measurements of display informativeness.

Required mental resources was quantified using normalized-and-scaled measures of subjective mental workload (NASA TLX) and physiological measures according to Equation 10. In instructional efficiency research, workload assessments have almost exclusively used retrospective subjective ratings (Fiore et al., 2006; Paas et al., 2003; Paas & van Merriënboer, 1993), which are appropriate for the longer time intervals involved in long-term memory encoding and skill development. In contrast, our interests concern short intervals, working memory processing, immediate application of task-relevant information, and continuous activity that ideally is not disrupted by workload surveys. Therefore, a major goal of the current research is to determine whether real-time physiological assessments can be as indicative of mental load as retrospective subjective ratings, if not moreso.

The results are negative with respect to this goal. Across all participants, the NASA TLX subjective ratings were more sensitive to display changes (see Table 14) and correlated more strongly with multitask performance than did any of the physiological assessment methods (see Table 16). A number of reasons may contribute to this finding. First, it is important to note that a broad range of factors (including but not limited to mental workload) affect autonomic arousal and corresponding physiological indicators; controlling for non-workload factors is difficult and/or expensive (O'Donnell & Eggemeier, 1986; Vidulich & Tsang, 2012). In an effort to make our methods broadly applicable, affordable, and minimally obtrusive to operators, the sensor technologies used were fairly simple. More sophisticated technologies may produce better physiological results.

The lack of sensitivity of physiological measures may also be impacted by ceiling effects in the indicators of mental workload which were not present in subjective assessments. This possibility is supported by initial evidence showing that near the cognitive redline, some physiological indicators show "plateau" effects not seen with subjective measures (Rodriguez Paras, 2015; Rodriguez Paras, Yang, Tippey, & Ferris, 2015), reducing the sensitivity of physiological measures at extreme workload levels. In our study, participants may approach/exceed their cognitive redlines since their multitask performance significantly dropped in the Spatial condition (Wickens, Hollands, Parasuraman & Banbury, 2012). Therefore, the plateau effect on physiological arousal (see Table 14) is likely to reduce the sensitivity of all physiological measures.

CE as a Property of Displays in Human-Machine Systems

An important feature of the new CE model (Equation 6) is that its component measures of informativeness and required resources are functionally dependent on human (h), display (d), and contextual (k) factors. This emphasizes that CE calculations are *system* evaluations, and varying any of those factors may have complex emergent effects on calculated values. It also implies that CE comparisons, for example, between hypothetical displays "d_A" and "d_B", are only meaningful when relevant human (h_i) and contextual/task (k_i) factors are held constant (e.g., CE(h_i,d_A,k_i) vs. CE(h_i,d_B,k_i)) or otherwise accounted for quantitatively in metric calculations. It remains a topic for future research to mathematically account for key human and contextual factors that impact CE measures so as to make this metric more generalizable.

CE Metrics for Predicting Multitask Performance

The current study illustrated the sensitivities of several CE metrics to Display conditions. Numeric, Hue, and Spatial displays were designed to be increasingly difficult to work with (communicating less information and/or imposing a greater cost to mental resources) and CE values decreased correspondingly, with INF/TLX distinguishing the displays statistically (see Table 15). The Spatial display in particular was designed to be especially demanding of visual-spatial working memory resources and create the most processing conflict between the LST and OAT. As expected, the Spatial display was consistently measured as the least cognitively efficient one and associated with the worst multitask performance.

Of the five constructed CE metrics (see Table 16), the INF/TLX metrics provided the highest correlation with multitask performance (PI) across all display conditions. This finding is consistent with the fact that NASA TLX is sensitive to the mental workload imposed by the LST, and validates the use of subjective measures in the previous instructional efficiency studies (e.g., Paas et al., 2003).

An unexpected finding was the degree to which interindividual differences influenced the sensitivity of CE metrics and their power to predict multitask performance. The High group demonstrated significantly higher cognitive efficiency than the Low group across all CE metrics (see Table 15), but the CE metrics are more predictive of multitask performance in the Low group (see Table 16). Unlike the High group, the Low group may not be able to obtain sufficient information from less efficient displays, thus resulting in apparent performance degradation.

Validation of the CE Metric

Equation 6 was formulated as a “benefits per costs” ratio to express the efficiency of a human-display system more completely than considering benefits or costs in isolation. CE quantifications can then meaningfully evaluate both highly-informative-but-complex and less-informative-yet-simple displays in a given task environment. For INF/TLX in the Low group, the ratio combination of measures in Equation 6 showed better predictive power for multitask performance than did component individual measures (see Table 16). These findings validate the ratio model and illustrate the predictive power of CE metrics in high-workload multitasking contexts.

Limitations

Although cognitive efficiency metrics were strong predictors of multitask performance in the current study, it must be emphasized that the results are meaningful only in the particular task context used in the experiment. Generalizing these methods and/or findings to other contexts requires additional consideration for different tasks and/or imposed loads. If the overall load does not approach an operator’s cognitive redline, CE measures may dissociate from task performance, because with excess mental resources available, operators can invest more (Wickens et. al., 2012) and thus performance with low-efficiency displays may be similar to that with high-efficiency displays.

Another limitation involves artificiality of the tasks; this was necessary to produce experimental conditions that facilitated the desired calculations. The displays were also very simple so that *informativeness* could be measured with practical methods. Accurately quantifying the CE dimensions with more complex real-world displays and tasks may require more sophisticated measurement techniques.

IV.3 Conclusion

In this chapter, the modified CE metric provides a powerful way to measure how well displays support multitask performance in human-machine systems. Versions of CE metrics that included NASA TLX subjective ratings for workload assessments correlated highly with performance, especially when operators were most heavily loaded, approaching their cognitive redlines. Importantly, some CE metrics showed higher correlations with performance indices than did constituent component measures, illustrating the value of the combined construct. With the ultimate goal of developing a practical method for evaluating display effectiveness in multitask contexts, the research in this chapter provides important theoretical basis for combating the problems of data overload and multitask performance breakdowns in a wide range of work environments.

CHAPTER V

INVESTIGATING A SPECIAL DISPLAY DIMENSION: BEAT PATTERN*

According to Chapter II, the selection of different display modalities and dimensions can largely affect the information perception and change the cognitive efficiency of display. In this chapter, we studied a special display dimension: beat pattern. Since beat pattern in auditory display (auditory beats) has been largely studied before, this chapter primarily investigated the beat pattern in haptic display (haptic beats), exploring its 'emergent property' and the effects of body location, sex, and beat frequency on the perception of haptic beats.

V.1 The Definition of Beat Pattern

A display can encode information into one or multiple of its dimensions. The diverse dimensions of visual, auditory, and haptic displays are summarized in Table 17.

Table 17: The Summary of Dimensions of Visual, Auditory, and Haptic Displays

Visual Dimensions: hue, brightness, size, shape, alphabetic, numeric
Auditory Dimensions: pitch/frequency, loudness, timbre
Haptic Dimensions: frequency, amplitude (intensity), waveform

* Reprinted with permission from *Proceedings of the Human Factors Society Annual Meeting*, 58, 1.

Table 17: Continued

General Dimensions of all three modalities: tempo, location, direction, duration, number

The beat pattern (not listed in the table) is a special dimension that can be found in the auditory and tactile presentations, being called ‘auditory beats’ and ‘haptic beats’ respectively.

V.1.1 Auditory Beats

Auditory “beats” are an emergent property of complex auditory presentations that can be observed when two pure tones of different frequencies (i.e., pitches) are presented simultaneously to the same observer (Oster, 1973). Beats can be observed both when the component auditory stimuli are integrated in a single auditory stream and when they are presented separately to the two ears. When the auditory stimuli are presented separately, they are integrated in the superior olivary nucleus and referred to as “binaural” beats (Oster, 1973).

The emergent perception of auditory beats arises from the formation of a complex tone that follows a combined sinusoidal pattern of amplitude modulation (see Figure 20). The frequency of the resultant sinusoidal pattern (i.e., the “beat frequency”) is equal to the difference between the signal frequencies of the component tones. For example, a simultaneous presentation of 1000 Hz and 1005 Hz signals would form a complex tone with a 5 Hz beat frequency.

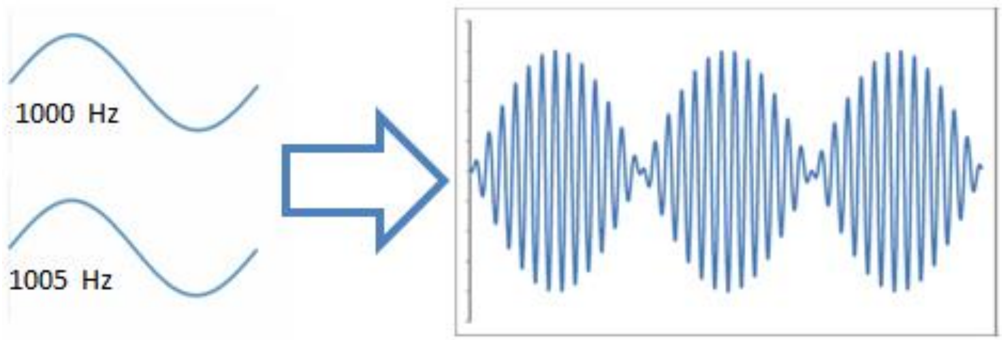


Figure 20. Graphical representation of amplitude-modulated “beats”, which is the combination of two sinusoidal waves.

Auditory beat frequencies slow and become imperceptible as the component frequencies approach the same value. This quality allows auditory beats to be used in calibrating devices emitting distinct auditory streams, such as when tuning a musical instrument or in ensuring jet engines are operating at the same frequency (Oster, 1973). Binaural beat presentations have also been used as an alternative method to treat depression because they increase the depth of meditation (Wahben, Calabrese & Zwickey, 2007). Auditory beats can also be used to encode data in sonifications and other synthetic auditory displays; for example, the deviation from a target driving speed can be encoded as the frequency of auditory beats (Yang, You & Ferris, 2013). Auditory beats can enhance the salience of signals in a noisy environment and, unlike pure auditory tones, the ability to detect beats does not deteriorate with age (Oster, 1973).

V.1.2 Haptic Beats

With a recent growth in the research and commercial development of *tactile* displays (e.g., Jones & Sarter, 2008;), haptic/vibrotactile beats are increasingly being explored as an additional display dimension to enhance the expressiveness of tactile displays (e.g., Ferris & Sarter, 2011; Hoggan & Brewster, 2007). While some vibrotactile display hardware has the ability to produce an integrated beat frequency signal within a single actuator, it has also been demonstrated that humans can perceive haptic beats when two distinct devices present pure vibrating signals to proximal but separate body locations. In this sense, the beat frequencies are an emergent property of the combined presentation.

The emergent perception of haptic beats has been demonstrated at the fingertip when holding two vibrating devices that are in direct contact (Lim, Kim, Hwang, & Kwon, 2010), when attaching separate vibrotactile devices to the same fingertip (Lim, Kyung & Kwon, 2012), and when touching a vibrating finger to either a mobile screen or to the skin while it is being vibrated by a separate actuator (Makino & Maeno, 2013). For haptic beats to be observed, the difference in activation frequency of the component vibrotactile signals must stay within the range of just noticeable difference (JND) (i.e., if presenting the two component vibrotactile signals individually, people would not be able to determine that the activation frequencies are different) (Pongrac, 2007).

The limited set of studies exploring the emergent perception of haptic beats from multiple stimulation sites has concentrated on the fingertip. However, other body locations may also support this perception. Promising sites include the palm and forearm, which can support the interpretation of spatial vibrotactile patterns with relatively high accuracy (Piateski & Jones, 2005), and the joints of the elbow and wrist, which support higher vibrotactile localization than other locations on the forearm (Chen, Santos, Graves, Kim & Tan, 2008; Cholewiak & Collins, 2003).

As prior research on emergent haptic beats involved presentation locations in very close proximity on or near the pad of the fingertip, the perception of beats is likely attributable to the physical propagation of vibration waves through the skin. This would mean the signals integrate locally via the mechanoreceptors in the skin to form a complex vibratory beat sensation. However, as in the perception of auditory binaural beats, these component vibrotactile waveforms may alternatively be sensed independently when at distinct body locations, with the sensory signals transmitted via parallel neural fibers to the cerebral cortex and integrated to form the perception of an emergent beat frequency. To gain insight into the underlying mechanisms in emergent haptic beats perception, the current study explored proximal presentation locations at the fingertip, palm, wrist, and elbow as well as distant body locations that do not support mechanical propagation and where emergent perception of beats would indicate neural mechanisms for integration of the vibrotactile stimuli.

V.2 A Case Study of the Perception of Haptic Beats

The goal of this study is to gain a deeper understanding of the mechanisms behind the perception of haptic beats and to identify where the emergent perception of beats is strongest. To achieve this goal, this study tested participants' ability to distinguish emergent haptic beats from pairs of pure vibration presentations as a function of body location and frequency difference. One grouping of body locations tested spatially-proximal paired presentation sites to examine how the local integration of component vibratory stimuli propagating through the skin may be affected by underlying tissue qualities (Group 1; see Table 18). Another grouping investigated whether haptic beats could be detected with paired presentation locations that were less proximal and involve different neural pathways, thus requiring beat perception to emerge from neural integration (Group 2; Table 18). In Group 2, paired presentations to fingertips on separate hands corresponds with detection across different cerebral hemispheres, and the use of two fingertips on the same hand and a fingertip and palm position on the same hand correspond with detection within parallel nerve fibers linked to one hemisphere. Additionally, sex effects were also evaluated for both location groupings, as previous studies of tactile perception found them significant (e.g., Komiyama, Kawara & De Laat, 2007).

Table 18: Grouping of Paired Presentation Locations

Group 1	(1) Fingertip: Pad and nail of the same middle finger
	(2) Wrist: Ulnar and radial styloid processes
	(3) Elbow: Lateral and medial epicondyle
	(4) Palm: Thenar and hypothenar of the same palm

Table 18: Continued

Group 2	(5) TwoFingers: Pad of the index finger and pad of the middle finger on the same hand (6) FingerPalm: Middle finger pad and hypothenar of the same hand (7) TwoHands: Pads of the middle fingers on two separate hands

V.2.1 Method

Twelve participants (6 males and 6 females) from Texas A&M University participated in this study. All participants were between 20 and 32 years of age (average age 25) and had no known injuries or conditions that would affect the ability to detect vibrations presented to the hand or lower arm.

Apparatus and Stimulus Presentations

The EAI© C-2 Tactor system was used to drive two vibrotactile actuators (referred to as "tactors") that presented the stimuli. Software was developed to simultaneously activate these tactors at either the same frequency or two slightly different frequencies. The tactors were attached to participants using compression fabric and Velcro at approximately equal pressures (see Figure 21). Participants were asked to assist in affixing the tactors to ensure satisfactory comfort levels. Participants wore noise cancelling headphones to minimize the audibility of tactor activation.

Vibrotactile stimuli were presented to seven pairs of body locations, which were divided into Group 1 and Group 2 (see Table 18).

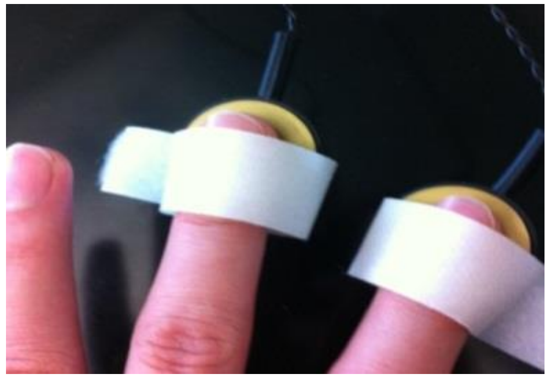


Figure 21. C-2 Tactors attached to the pad of middle and index fingers (“TwoFingers” – see Table 18).

At each location, two different experimental frequency pairings with the potential to produce haptic beats were tested: (250 Hz from tactor 1, 246 Hz from tactor 2) – with a resultant beat frequency of 4 Hz – and (250 Hz, 242 Hz) – with a resultant beat frequency of 8 Hz. There were 14 total (7 location pairings x 2 frequency differences) treatment conditions, each of which required reporting whether or not haptic beats were observed for ten successive presentations. The ten trials consisted of a random ordering of five presentations with the specified frequency difference and five control trials in which the vibrations were the same (250 Hz) at both locations. A base frequency of 250 Hz was chosen because it is within the range of maximum sensitivity for Pacinian corpuscles, which play a large role in vibrotactile sensation in the hands and joints (Cholewiak & Collins, 1991; Sherrick & Cholewiak, 1986), and it is also near the mechanically optimal operational frequency for the tactor hardware. The deviations (4 Hz and 8 Hz) from the “carrier” frequency (250 Hz) were within the just noticeable difference (JND) range

(Pongrac, 2007) and were similar to differences used in previous studies of emergent beats perception at the fingertip (Lim, Kyung & Kwon, 2012). The intensities of vibration presentations were comparable to a cell phone set to vibrate mode, with the 7.5-mm “skin contactor” of the tactors oscillating at a maximum displacement of approximately 1 mm.

The order of location presentations was balanced within each Group, but Group 1 was always presented before Group 2. The order of resultant beat frequencies and the presentations of the ten stimuli in each location-frequency pair treatment condition were quasi-randomly distributed.

Procedure

After consenting to participate and completing a background questionnaire, participants were introduced to the haptic beats concept and experienced the beat sensation at the 4 Hz and 8 Hz beat frequencies. When presented to the fingertip, all participants were able to distinguish a pure vibration presentation from a beat presentation with 100% accuracy, affirming that they clearly understood the haptic beats concept. Participants then completed all 14 treatment conditions, for a total of 140 trials. For each trial, participants were simply required to verbally report “Yes” if they felt a beat frequency or “No” if they did not feel a beat frequency (i.e., if they felt a pure signal). Prior to the start of each new treatment condition, participants were trained on a set of at least 6 randomly ordered stimuli for which they were given feedback about the correctness of their response; this procedure was repeated until their recognition of beat frequencies appeared to reach an asymptote. Each trial presentation lasted approximately 2.5 seconds with 4 seconds between trials. There was one minute of required rest time between treatment conditions. In total, the experiment lasted approximately one hour.

V.2.2 Results

The accuracy of response was recorded as the percentage of trials in which the participants correctly identified either haptic beats or pure signal presentations. Repeated Measures Three-Way ANOVAs formulated in Minitab were used to examine the treatment effects of body locations and beat frequencies on the perception of haptic beats for each location grouping.

Results of Group 1

The fingertip, palm, wrist and elbow locations showed overall identification accuracies of 98.8%, 90.4%, 91.3% and 93.3%, respectively. The accuracies differed significantly among these four body locations (F (3, 80) = 3.30; p = 0.025). Post-hoc Tukey tests showed significantly higher accuracies for the fingertip compared to palm (p=0.0276). The accuracies for the fingertip, wrist, and elbow did not significantly differ from each other.

The accuracies for conditions involving 8 Hz beat frequency presentations (mean accuracy: 97.3%) were significantly higher than conditions involving 4 Hz presentations (mean accuracy: 89.6%; $F(1, 80) = 13.94$; $p < 0.001$). Male participants also responded with significantly higher accuracies (mean accuracy: 96.9%) than females did (mean accuracy: 90.0%; $F(1, 80) = 11.09$; $p = 0.001$). None of the two- and three-way interaction effects between body location, frequency, and sex were significant.

Although no significant interaction effect was found, there was a prior expectation that the difference in salience between the 4 Hz and 8 Hz beats presentations could affect beats perception at body locations with relatively lower sensitivity. Therefore we compared the accuracies for 4 Hz and 8 Hz beats at each body location (see Figure 22). Beat frequency did not significantly impact identification accuracies for presentations on the fingertip, with 98.3% accuracy for 4 Hz beats and 99.2% accuracy for 8 Hz beats. For the palm, the average accuracy for 4 Hz treatment conditions (mean accuracy: 85.8%) is much lower than that for 8 Hz conditions (mean accuracy: 95.0%), although this difference did not reach significance. However, the wrist ($F(1, 20) = 5.49$; $p = 0.030$) and elbow ($F(1, 20) = 6.74$; $p = 0.017$) demonstrate a significantly greater accuracy for treatment conditions involving 8 Hz haptic beats compared to those involving 4Hz beats (Wrist: 4 Hz = 85.0%, 8 Hz = 97.5%; Elbow: 4 Hz = 89.2%, 8 Hz = 97.5%).

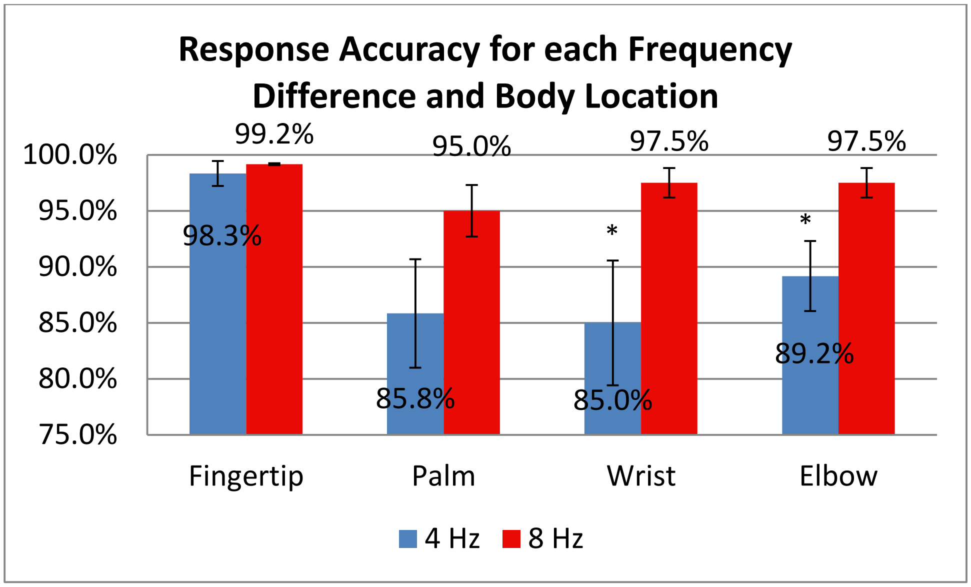


Figure 22. Response accuracy for each frequency difference and body location in Group 1. * indicates significance, and error bars represent standard error.

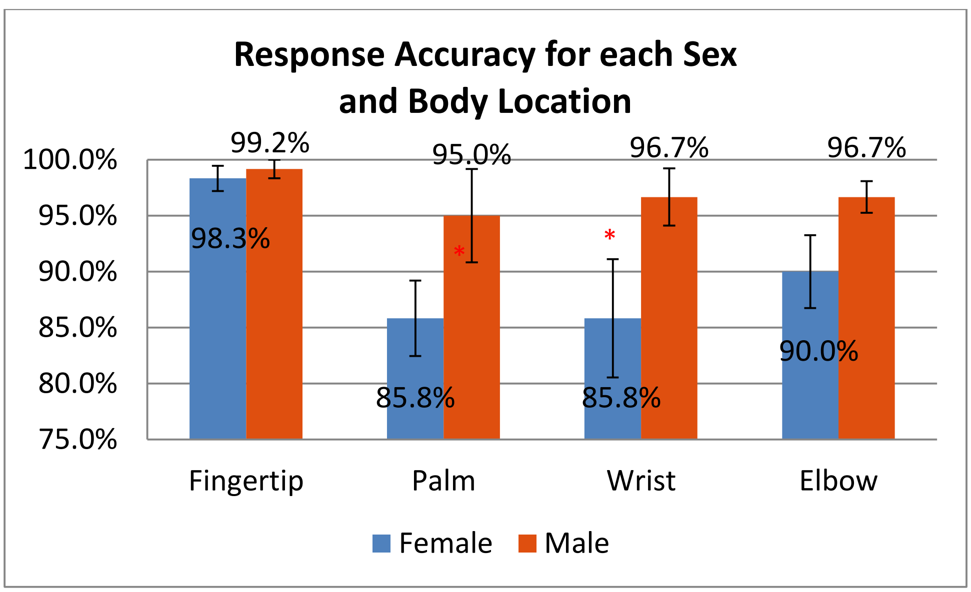


Figure 23. Response accuracy for each sex and body location in Group 1. * indicates marginal significance, and error bars represent standard error.

Because previous studies had found consistent sex differences in auditory binaural beats perception (Tobias, 1965; McFadden, 1998), the sex effect for haptic beats perception was also analyzed at each body location (see Figure 23). Males and females accurately identified 99.2% and 98.3% of the stimuli, respectively, with their fingertips. For the other body locations, males' average identification accuracy (Palm = 95.0%, Wrist = 96.7%, Elbow = 96.7%) was higher than that of females (Palm = 85.8%, Wrist = 85.8%, Elbow = 90.0%). None of these locations significantly differed due to sex, but the p-values of wrist ($F(1, 20) = 4.12$; $p = 0.056$) and elbow ($F(1, 20) = 4.32$; $p = 0.051$) are very close to the chosen significance criterion ($p = 0.05$).

Results of Group 2

None of the treatment conditions in Group 2 showed accuracies that were significantly different from chance (mean accuracy: TwoHands = 48.3%, TwoFingers = 52.9%, FingerPalm = 56.7%). The effects of frequency and sex were also insignificant.

V.2.3 Discussion

Haptic beats offer a promising means of increasing the expressiveness of haptic displays and supporting the presentation of more complex signals. This phenomenon can be observed when vibrotactile stimuli of different frequencies are presented simultaneously to one or multiple body locations. Recent research has demonstrated the emergent perception of beats for paired presentations to the pad and nail of the same fingertip but has left other body locations and the effects of frequency ranges largely unexplored. In order to further the understanding of haptic beat perception, its underlying neural basis, and its potential applications, this study aimed to measure the ability to reliably detect haptic beats at seven pairs of locations on the lower arm and at two beat frequencies.

Effect of Location

Consistent with previous findings, the fingertip demonstrated a very strong ability to detect emergent haptic beats that did not differ due to frequency or sex effects. This corresponds with the fingertip's relatively high sensitivity for processing vibratory stimuli. This body location contains a high innervation density for key types of mechanoreceptors, especially fast adapting type II (FA II) receptors that are easily excited by mechanical oscillations with frequencies ranging between 100 Hz and 300 Hz (Vallbo & Johansson, 1984). Furthermore, the cortical representation of fingers occupies a large area of the somatosensory cortex (SI) (Penfield & Rasmussen, 1951), suggesting a high capacity for processing tactile information at the fingertip. Finally, the relative proximity of the locations on either side of the fingertip and the rigidity of the tissue in the fingertip support vibration propagation and localized mechanical integration of the paired waveforms. Therefore, the fingertip is likely the best body location, of those explored in this study, for perceiving complex patterns such as emergent haptic beats.

Beyond the fingertip, the palm, wrist, and elbow also displayed a strong ability to detect haptic beats. At these locations, anatomical differences between participants may have affected the results to some extent. For example, some participants appeared to have more underlying muscle and tendon tissue on the elbow and wrist. These participants tended to be less capable to detect haptic beats at these locations. This observation suggests that more direct stimulation of bone enhances the sensitivity to haptic beats since rigid bone tissue effectively propagates high-frequency physical oscillations (Bacabac, et.al, 2006). Taking advantage of the physical propagation of vibration signals within the bone at the wrist and elbow may be a reliable way to present emergent haptic beats, in a manner similar to how auditory beats can be presented using binaural bone conduction in auditory displays (Walker, Stanley, Lyer, Simpson & Brungart, 2005).

This study also addressed the extent to which emergent haptic beats perception was attributable to localized mechanisms, such as pairs of vibrations propagating through the skin, versus to neural mechanisms, such as parallel streams of neural signals being integrated at the cerebral cortex. Therefore, two groups of paired body locations tested each of these mechanisms, with those in Group 2 involving locations associated with different nerve fibers: different fingertips of the same hand, the palm and one of its connected fingertips, and corresponding fingertips of opposite hands. However, none of the Group 2 presentation locations enabled participants to distinguish vibrotactile stimuli at a rate significantly different from chance (50%). This finding suggests that haptic beats perception likely does not occur due to integration in the somatosensory cortex in any manner similar to the way signals are integrated into binaural beats in the auditory cortex, whether the component vibration signals are processed bilaterally (one component in each cerebral hemisphere, as in presentations to both hands) or unilaterally (as in presentations to different locations on the same hand). Thus, the more likely mechanism for emergent haptic beat perception is localized mechanical integration.

Effect of Frequency

The results also suggest that differences in beat frequencies affect the perception of haptic beats. Between the two frequency pairs, the larger difference (the 8 Hz beat frequency) resulted in greater beat identifications accuracy, especially at the wrist and elbow, which have relatively lower sensitivities to complex/high frequency vibration in comparison to the hands and fingers (e.g., Penfield & Rasmussen, 1951). The higher 8 Hz beat frequency was also commonly reported as feeling more salient than the 4 Hz frequency, which may have contributed to this difference. However, the potential for improved beat detection from continuing to increase the beat frequency within the range of Just Noticeable Difference is unclear, suggesting a direction for future study. As one common mapping for haptic beat frequencies is the urgency of an encoded message, such as in displays communicating the severity of deviation for monitored patient health parameters (Ferris & Sarter, 2011) and the urgency in which changes in a driver's speed are needed to achieve a target speed (Yang, You & Ferris, 2013), more urgent messages should be encoded with higher beat frequencies, which are more salient and thus more likely to be detected.

Effect of Sex

A sex difference also exists in the perception of haptic beats, with males' ability to detect beats appearing stronger than females'. As with the differences due to beat frequency, this difference was more pronounced at the less-sensitive locations of Group 1: the wrist and elbow. Although sex differences have been observed in tactile detection thresholds at some body locations, such as on the skin of the cheek (Komiyama, et al., 2007), sufficient evidence does not exist to determine if a true detection difference exists on the forearm due to the anatomical differences within the study population. Consistent with data on average anthropometries, female participants in this study were observed to have relatively smaller bones and less flat surface area at the wrist and elbow than male participants. This made it difficult to provide reliable skin contact for females at these body locations with the C-2 tactors and may have been the primary reason for the observed sex differences. Future studies are needed to explore anatomical differences and other possible factors contributing to the difference in sex perception, as well as to include larger sample sizes to more precisely control for these factors.

Emergent Property

Haptic beats offer a promising "emergent property" to employ in advanced display methods such as configural and object displays (Bennett & Flach, 1992). For example, if levels of variables of interest are encoded into the frequency dimension of a vibrotactile display, then the emergent perception of haptic beats can communicate a difference in levels among related variables. For a human operator monitoring a set of automated subsystems, a complex vibrotactile display that includes several proximal presentation locations – each representing a single subsystem – could use a common frequency to communicate nominal conditions for each subsystem. Deviation from the nominal condition in at least one subsystem would result in a frequency change at the corresponding display location. The effective perception would then morph from a pure vibratory signal across the entire display area to a salient emergent beat. This would allow a holistic view of overall system performance for the operator as well as directing attention to sources of the deviation when they occur. Additional applications include the use of haptic beats in calibration tasks, such as those currently performed with auditory beats; in remote target-approach or guidance tasks, such as indicating the proximity of the position of a robot arm to a task location; and in personal biofeedback devices, such as informing wearers with a heart condition that their heart rate is approaching an unsafe level.

Limitations

Some limitations of this study include a limited sample size, the potential for vibrotactile propagation through the compression fabric for some locations, and uncontrolled skin temperature, which has the potential to impact peripheral tactile sensitivity (Green, 1977). These limitations, as well as sex differences, will be addressed and further explored in future studies.

V.3 Conclusion

Chapter V investigated an important dimension of display: beat pattern. We studied haptic beats which is the beat pattern of haptic display and found that haptic beats can be reliably perceived at the fingertip, palm, wrist and elbow, and likely any spatially-proximal paired locations that support the localized integration of vibrations which physically propagate through tissue. Frequency differences of the component signals and sex also significantly affect the perception of haptic beats, especially at locations with lower tactile sensitivity such as at the wrist and elbow joints.

Consistent with previous findings, haptic beats were reliably perceived with paired presentations on the same fingertip; previously-unexplored locations on the palm, wrist, and elbow also supported perception of beats. However, haptic beats were not perceived when stimuli were presented to distant locations, such as on different hands, suggesting that haptic beats most likely involve a localized mechanical integration rather than neural integration.

The study of beat pattern can inform the design of advanced displays, such as configural or object displays, which can support direct perception of relationships among basic display components through "emergent properties" that arise from the arrangement of those components (e.g., Bennett & Flach, 1992). While such advanced display design efforts have focused primarily on visual displays, extending this concept to complex auditory and haptic displays can benefit domains that heavily load visual resources, such as driving, dismounted navigation, medicine, and immersive virtual reality.

CHAPTER VI

MEASURING THE COGNITIVE EFFICIENCY OF NOVEL SPEEDOMETER DISPLAYS*

The studies in previous chapters only measured the cognitive efficiency of visual display with discrete presentations; but in Chapter VI, we extended the application of CE metric to the continuous speedometer displays that engaged different perceptual modalities and dimensions This chapter evaluated the cognitive efficiency of an ambient-visual, an auditory, and a tactile display in multitask driving scenarios. Moreover, it investigated whether using beat pattern (studied in Chapter V) as a redundant dimension to encode information can improve the cognitive efficiency of the auditory and tactile displays. The findings of Chapter VI have implications for the design of efficient continuous displays for various work domains, such as driving, navigation, medicine, and immersive virtual reality.

VI.1 Measuring the Cognitive Efficiency of Novel Speedometer Displays

The goal of Chapter VI is to evaluate the cognitive efficiency of novel speedometer displays and find the most cognitively efficient modality and dimension of the displays in multitask environments. Therefore, this chapter answers the fourth research question of

---

the dissertation: **how is the cognitive efficiency of continuous speedometer displays?** This question is decomposed into two sub-questions, which are

1) **how is the efficiency of the speedometer displays that engaged different modalities in multitask settings**;

2) **do the displays that encode information into redundant dimensions are more cognitively efficient in multitask environments than the single-dimension displays**.

To answer the two questions, the study adopted a dual task set: display processing task (i.e., verbally report the target information presented by displays) and lane tracking task (i.e., maintain the vehicle position on the center of the line). The cognitive efficiency of display was measured based on the verbal report in the display processing task and the measures of imposed mental workload.

VI.1.1 Method

Participants

The study recruited 24 adults (12 males and 12 females) from Texas A&M University, with an average age of 27 years old. All students were at least 18 years old, had normal or corrected-to-normal visual and auditory acuity, and no known conditions that affect tactile perception on lower back. Each participant had a valid driving license for at least one year.

Displays

There were five types of displays - ambient-visual (AV), auditory-spatial (AS), auditory-redundant (AR), tactile-spatial (TS), and tactile-redundant (TR) - which were modified from the redundant displays used in the previous research for supporting speed tracking task (Yang et al., 2013; Yang et al., 2015). More technical details of building these displays can be found in these two studies.

The ambient-visual (AV) display consisted of a sequence of colors which were projected on a drop-down screen behind the computer monitor. The center point of the color sequence is yellow.

The auditory-spatial (AS) display consisted of active noise-cancelling headphones playing sounds from nine virtual origins which were created by SLAB (http://human-factors.arc.nasa.gov/slab/). The nine virtual locations were equally distributed as if originating from a semicircle on the transverse plane at the height of ear (see Figure 24). The radius of the semicircle was 50 cm. To enhance the perception of the virtual spatial location, the pitch of each auditory presentation was adjusted into distinguished levels which increased from the auditory presentation No.1 to No.9 (see Table 19).

The auditory-redundant (AR) display was modified from the auditory-spatial display by adding redundant beat patterns. When the speeds moved away from the center point towards either extreme, the beat frequency (i.e., the difference between paired frequencies) discretely increased to higher levels (see the numbers in the parentheses in Table 19). The center point of both auditory displays was denoted by sounds No.5 (See

Figure 24). Both auditory presentations were displayed at volumes that could be reliably and comfortably heard.

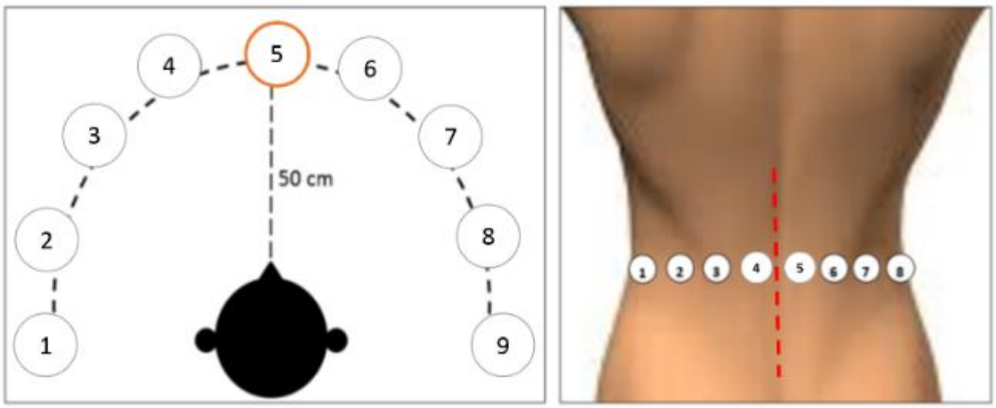


Figure 24. The nine virtual auditory locations surrounding the head (left) and the eight tactor locations on the lower back (right).

The tactile-spatial (TS) display incorporated eight C-2 tactors developed by Engineering Acoustics, Inc. The tactors were arranged horizontally across the lower back and symmetrically distributed on both sides of the spine. Tactors were doubly-secured with an elastic weightlifting belt and a strap over the top of the belt. In order to assure sensation but avoid annoyance, the tactors were presented at the medium gain. For the tactile-redundant (TR) display, the frequency of the beats increased as the vibration presentation site moved laterally away from the location of spine (see the numbers in the parentheses in Table 19).

All displays presented information as a sequence of *transition patterns*. In our definition, a transition pattern is a sequence of presentations that transform from the first one, across all presentations between, to the last along one direction (left or right). For example, the red color eventually transfers to the blue color, crossing each color between red and blue. The beginning and end of a transition pattern can be any of the presentations, with at least one presentation between them. The last presentation of each transition pattern overlaps with the first one of the next transition pattern, creating a sequence of consecutive transition patterns, which was used to simulate the pattern of speed fluctuation as showed in the previous studies (Yang et al., 2013; Yang et al., 2015).

Table 19: The Nine Presentations of each Display and Their Locations and Distances to the Center

Location	Left				Center	Right			
Distance	4	3	2	1	0	1	2	3	4
AV	(153,0,0)	(255,0,0)	(255,102,0)	(255,204,0)	(255,255,102)	(187,239,57)	(0,160,0)	(0,102,255)	(0,0,255)
AS	1; 690	2;587	3;493	4;492	5;349	6;329	7;261	8;220	9;174
AR	1;690+702	2;587+595	3;493+497	4;392+393	5;349	6;329+328	7;261+257	8;220+212	9;174+162
	(12)	(8)	(4)	(1)	(0)	(1)	(4)	(8)	(12)
TS	1;238	2;242	3;246	4;249	Spine	5;251	6;254	7;258	8;262
TR	1;238+250	2;242+250	3;246+250	4;249+250	Spine	5;251+250	6;254+250	7;258+250	8;262+250
	(12)	(8)	(4)	(1)		(1)	(4)	(8)	(12)

For all displays, the duration of a stimulus (SD) was 750 milliseconds. But their inter-stimuli interval (ISI) and the inter-movement interval (IMI) slightly varied. There were 5 presentations in each transition pattern on average for all displays. Each trial consisted of average 5 transition patterns. See Table 20.

Table 20: The Parameters of Displays which Engaged Different Modalities

Display	SD (ms)	ISI (ms)	IMI (ms)	Average Number of stimuli per transition	Average Number of transitions per trial
Visual	750	0	0	5.5	5
Auditory	750	50	100	5.4	5
Tactile	750	50	0	5.2	5

Mental Workload Measure

Three physiological measurement devices (i.e.,, Shimmer3, Bioharness3, and Pupil Labs) and one subjective measure (i.e., NASA TLX) were used to measure mental workload. However, only NASA TLX (Hart & Staveland, 1988) will be reported here. NASA TLX measures mental workload by asking participants to subjectively report their ratings of six scales and weight the scales in the end.

Task

Participants sat in the driving simulator and interacted with the driving scenarios created by STISIM Drive® Software via the Logitech G27 steering wheel and pedals. They drove on a one-way lane that consisted of curves one after another with varying types of curvatures. The vehicle speed was governed at constant 50 MPH in driving.

There were 6 task zones in each driving condition. Each task zone included two concurrent tasks: lane tracking and display processing. For the lane tracking task, we attached a foam tape to the desktop screen to indicate the center of vehicle. Participants needed to control the vehicle to overlap the foam tape with the center line. Half of the participants drove in windy conditions in which the wind (generated by complex sinusoid waves) continuously blew the vehicle back and forth along the lateral direction. The other half drove in normal conditions without the wind effect. The maximal lateral distance moved by wind was round 3 feet.

At the same time of lane tracking, participants needed to process a trail of information from one of the five displays. At the end of a task zone, the display ended and the driving scenario paused for 20 seconds during which the experimenter asked four questions (see Table 21) and participants verbally answered them after each question. Since it's difficult to calculate the information transmitted of continuous information, we proposed informativenesss index (INF) to indicate the overall informativeness of each display. It was calculated as the sum of the score of each question (see Table 21).

Table 21: The Score of each Question

Question Description	Response		Score
Is the last presentation of the trial on the right side or left side?	left right	S1:	1, f correct 0, if wrong
What's the direction of the last transition pattern, left-to-right or opposite?	left right	S2:	1, if correct 0, if wrong
What's the distance between the last presentation of the trial and the center point?	1, 2 ,3 ,4	S3:	2- |error|*0.5
How many transition patterns pass the center point?	1 ~ 7	S4:	2.8, if error=0 1.4, if error=1 0.7, if error=2 0, if error>2

Note. Error is the absolute difference between response and correct answer.

Procedure

After signing the consent forms and finishing the background questionnaires, participants were asked to record a 5-minute physiological baseline. Then, experimenter used PowerPoint (with animation) to introduce each type of display. After the introduction of each display (except ambient-visual), participants need to be familiar with the presentations by identifying the random presentations until they can't achieve higher accuracy. Then they experienced each type of display processing task in the training scenarios and understood what they should process and report.

After the training, participants went through six experimental conditions (one baseline and five display conditions) in a balanced order. Before each experimental condition, experimenters repeated the nine presentations of the respective display. Each experimental condition took around 3 minutes to complete. During the interval between consecutive conditions, participants completed a NASA TLX rating questionnaire and took a 90-second breathing exercise to recover the physiological baseline. Finally, participants completed a weighting questionnaire of NASA TLX.

Data Analysis

*D*riving performance was measured by lane deviation. Display informativeness was indicated by informativeness index (INF) and mental workload engaged by display was calculated as TLXDisplay (TLXDisplay = TLX-TLXBaseline). Both INF and TLXDsiplay were normalized and scaled to the same positive range and then INF was divided by TLXDisplay (INF/TLXDisplay) to indicate cognitive efficiency. All dependent variables were analyzed in the three-way repeated measures ANOVAs (Experimental Condition × Wind × Sex) in R 3.1.3. Tukey post-hoc tests were applied for pairwise comparisons.

VI.1.2. Results

Lane Deviation (LD)

LD was not significantly different among the experimental conditions. However, the main effects of sex ($F(1,120)=64.04$, $p<0.001$, $\eta^2=0.415$) and wind ($F(1,120)=17.13$, $p<0.001$, $\eta^2=0.111$) were significant on LD. LD in windy condition (LD=2.00) was significantly larger than that in normal condition (LD=1.31) and male showed much lower LD (0.99) than female (2.33).

Moreover, the interaction effect between sex and wind was also significant on LD ($F(1,120)=39.43$, $p<0.001$, $\eta^2=0.112$). Female participants had significant lower LD than male in either windy (LD: female=3.11, male=0.99, $p<0.001$) or normal (LD: female=1.64, male=0.99, $p<0.001$) condition. Female participants showed significantly higher LD in the windy condition comparing to the normal condition ($p<0.001$). See Figure 25.

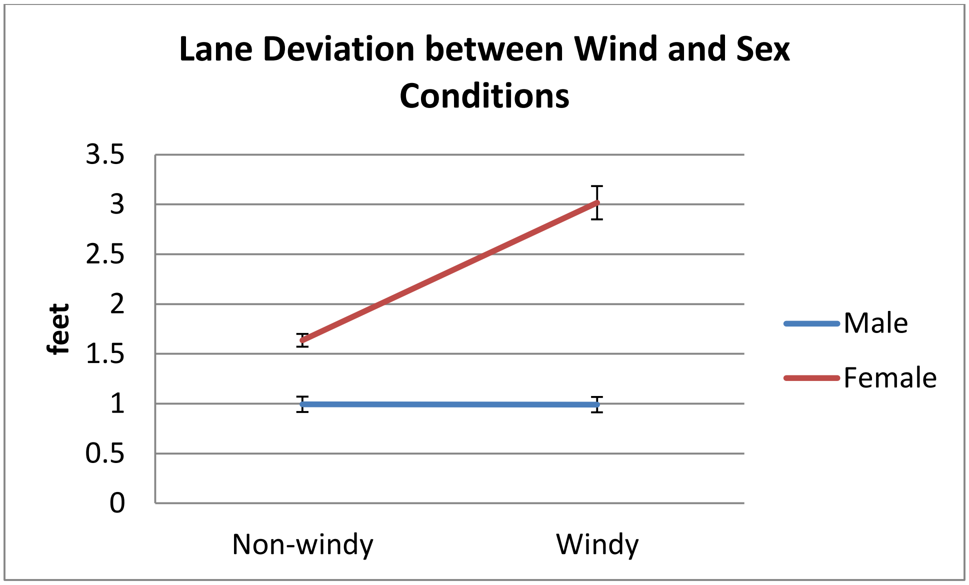


Figure 25. Lane deviation between wind and sex conditions.

Informativeness Index (INF)

INF was significantly different among display conditions ($F(4,100)=9.17$, $p<0.001$, $\eta^2=0.257$). Post hoc test showed that the auditory-spatial display (INF=4.83) had significantly lower INF than that of the auditory-redundant (INF=5.50, $p=0.006$), tactile-spatial (INF=5.79, $p<0.011$), and tactile-redundant (INF=5.45, $p<0.001$) displays. Similarly, the INF of ambient-visual display (INF=4.93) was also significantly lower than that of the auditory-redundant ($p=0.027$), tactile-spatial ($p<0.001$), and tactile-redundant ($p=0.049$) displays. See Figure 26. None of other effects on INF was significant.

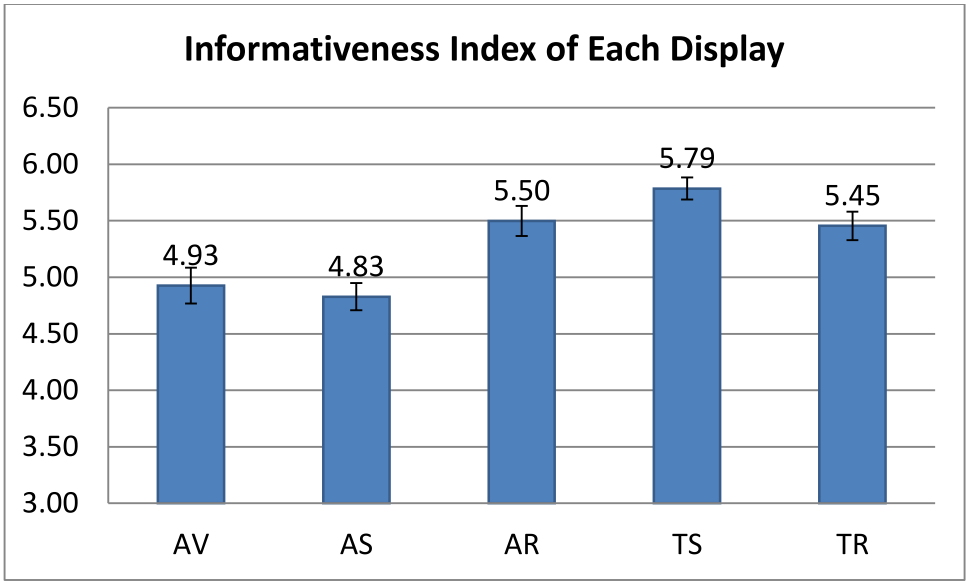


Figure 26. The informativeness index of each novel speedometer display.

NASA TLX

The effects of experimental condition (F(5,120)=13.87, p<0.001, η^2=0.315) and sex (F(1,120)=5.15, p=0.025, η^2=0.023) were significant on NASA TLX. Post hoc test showed that Baseline condition had significantly lower NASA TLX (80.0) than other conditions (NASA TLX: ambient-visual=150.4, auditory-spatial=185.7, auditory-redundant=158.0, tactile-spatial=149.8, tactile-redundant= 149.6, p<0.001). See Figure 27. Moreover, female participants reported significantly higher NASA TLX (154.26) than males (136.89).

After using baseline as covariate for baseline correction, it's found that auditory-spatial display had significantly higher NASA TLX than ambient-visual (p=0.047), tactile spatial (p=0.041), tactile-redundant (p=0.039).

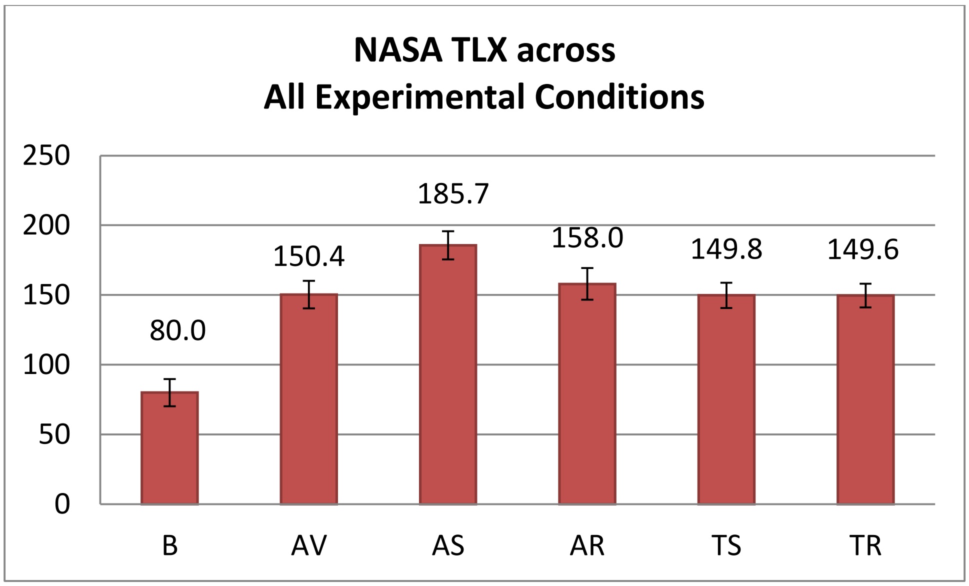


Figure 27. NASA TLX across all experimental conditions.

The interaction effect between sex and wind ($F(1,120)=18.39$, $p<0.001$, $\eta^2=0.084$) was significant on NASA TLX. For male participants, NASA TLX was significantly lower in the normal condition (114.3) than windy condition (159.5, $p<0.001$). In the normal condition, male participants had significantly lower NASA TLX (114.3) than female participants (164.5, $p<0.001$). See Figure 28.

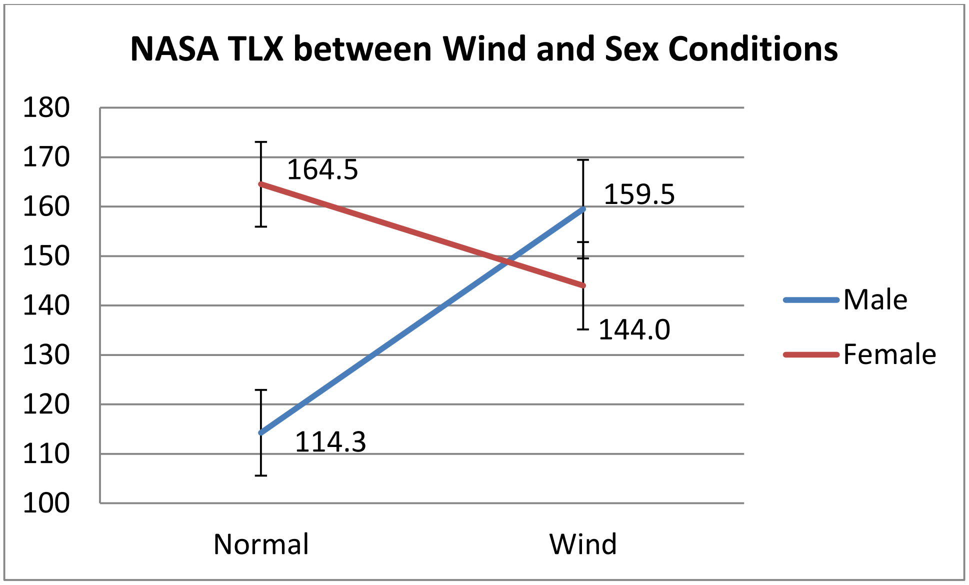


Figure 28. NASA TLX between wind and sex conditions.

Cognitive Efficiency (CE)

The main effects of display condition ($F(4,100)=86.07$, $p<0.001$, $\eta^2=0.173$) and wind ($F(1,120)=5.03$, $p=0.027$, $\eta^2=0.036$) were significant on the CE of displays. Post hoc tests showed the CE of auditory-spatial display was significantly lower than the auditory-redundant ($p=0.003$), tactile-spatial ($p<0.001$), and tactile-redundant ($p=0.006$) displays. See Figure 29. CE in windy condition (CE=0.87) was lower than that in normal condition (CE=0.97).

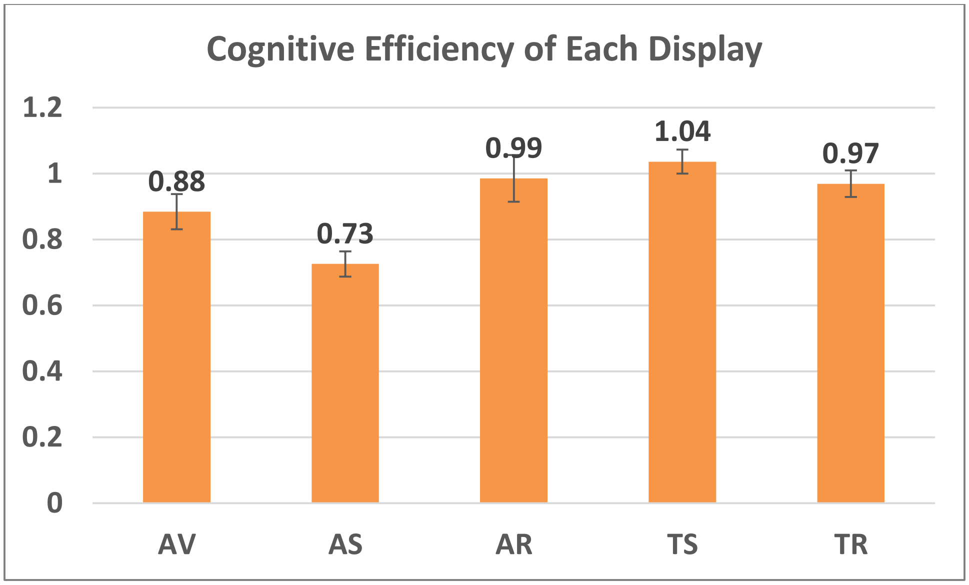


Figure 29. The cognitive efficiency of each novel speedometer display.

VI.1.3 Discussion

Chapter VI applied Cognitive Efficiency (CE) metric to assess the novel speedometer displays which engaged different perceptual modalities and dimensions. As a systematic approach, CE metric integrates the measures of display informativeness and imposed mental workload and provides a multi-factor evaluation of displays in human-machine systems. The measure of cognitive efficiency of displays will be used to predicate multitask performance, based on the finding of their positive correlation in Chapter IV.

Lane Deviation

The lane tracking task elevated the baseline of mental workload. The results showed that adding an additional display processing task didn't change the lane tracking performance, which suggested that participants distributed similar amount of attention to support the lane tracking performance in all display conditions.

CE of Displays with Different Modalities

The cognitive efficiency of the ambient-visual, auditory, and tactile displays (except the auditory-spatial display) in multitasking environments were at the same level. However, it is interesting to notice that the ambient-visual display was less informative as the auditory and tactile displays. Perhaps, the peripheral vision on the ambient-visual display and focal vision on the road interrupted each other since the two types of visions are not as distinctly separable as visual vs. auditory or visual vs. tactile modalities. The fact that the ambient-visual display was less informative but produced better cognitive efficiency should be the result of fewer imposed mental workload, although the low mental workload was not clearly showed by NASA TLX.

CE of Displays with Different Dimensions

In addition to modalities, cognitive efficiency of a display also depends on which display dimensions are chosen to encode information. Choosing the 'wrong' display dimensions can harm the effectiveness and efficiency of a display. For example, the auditory-spatial display was less informative and cognitively efficient than other displays, because human has limited capacity to recognize the auditory locations and absolute pitches.

Between the auditory displays, the auditory-redundant display which engages additional beat pattern was much more informative and required less mental resources to process, comparing to the auditory-spatial display (without beat pattern). This finding can be explained by fact that humans are naturally good at recognizing auditory beats (Oster, 1973). Therefore, beat pattern can be used as a promising feature for auditory displays to enhance its cognitive efficiency.

Between the two tactile displays, the tactile-redundant display (which engaged additional haptic beats) showed the same level of informativeness, mental workload, and cognitive efficiency as the tactile-spatial display. In other words, compared to spatial location, beat pattern may be a less effective haptic dimension to encode continuous information.

Wind Effect on CE

Contextual factors also significantly affected the cognitive efficiency of display. The results showed that the cognitive efficiency of all displays was lower in the windy condition than the normal condition. But informativeness index and NASA TLX were not significantly different between these two conditions. Maybe cognitive efficiency is more sensitive to contextual factors than its constituent components (display informativeness and imposed mental workload). Further studies are needed to explore more contextual impacts on cognitive efficiency and each of its constituent components.

Sex Effect

The informative index, imposed mental workload (total workload – baseline workload of lane tracking task), and cognitive efficiency of the same display were not different between the male and female participants, which meant both males and females did well in the display processing task. However, the female participants performed much worse in the lane tracking task than the male participants, especially in the windy condition. This was consistent with the observation in the experiment that the female participants were relatively weak at manipulating the steer wheel to control the lateral position of the vehicle. Thus, sex factor needs to be taken into account in the design of human-machine systems that require physical task performance.

VI.2 Conclusion

This chapter evaluated the cognitive efficiency of five novel speedometer displays which engaged different perceptual modalities and dimensions. The results showed that the selection of engaged modality can significant affect the cognitive efficiency of a display, but the effect also highly depends on the engaged dimensions of the modality. Moreover, encoding information into beat dimension can significantly improve the cognitive efficiency of the auditory display, but not of the tactile display. In addition, cognitive efficiency was found to be more sensitive to contextual factors (e.g., wind effect) than each of its constituent components. Built upon these findings, the next chapter will move forward to investigate how the novel speedometer displays support drivers' continuously tracking performance and examine the relationship between cognitive efficiency and tracking performance.

CHAPTER VII

SUPPORTING CONCURRENT TRACKING PERFORMANCE USING NOVEL SPEEDOMETER DISPLAYS*

The previous chapter evaluated the cognitive efficiency of five novel speedometer displays. In Chapter VII, we investigated the effects of these five displays on concurrent tracking performance which is important in the operations of current human-machine systems. In the driving simulation study of this chapter, participants can obtain information from a redundant novel speedometer display to support the speed tracking performance. At the same time, they also needed to conduct a lane tracking task (which was to keep the vehicle at the center of the lane). The findings of this chapter examined whether the continuous novel speedometer displays that showed higher cognitive efficiency in previous chapter can better support concurrent tracking performance. Moreover, this chapter provided valuable suggestions for the design of displays that support concurrent tracking performance in driving and many other task scenarios.

---

* Part of this chapter is reprinted with permission from *Proceedings of the Human Factors Society Annual Meeting*, 57, 1.
* Part of this chapter is reprinted with permission from *Proceedings of the Human Factors Society Annual Meeting*, 59, 1.

VII.1 Concurrent Tracking Performance

According to the National Highway Traffic Safety Administration Traffic Safety Facts, speeding contributes to approximately 30 percent of the total fatalities. Therefore, speed management is critical for driving safety. However, managing speed may interfere with other concurrent driving tasks which are also important for driving safety. For example, the work and school zones require drivers to control the speed carefully at an unnaturally slow level while keeping their eyes closely on the road. But these two tasks compete for the limited visual attentional resources since sampling speed information from the speedometer moves eyes away from the road. This attentional conflict in driving illustrates the need of supporting speed management in visual-demanding environments.

The study in this chapter focuses on a specific type of speed management: speed tracking. A tracking task is defined as a task that requires operators to frequently monitor a task-related variable (e.g., speed or lane position), determine the perceived error between each variable's current value versus its target levels, and 'correct' that error. Tracking more than one variable at the same time generates *concurrent tracking task* which requires an operator to switch visual attention quickly among different target variables and modify their states, which can impose heavy mental workload on the operator and interrupt task performance. Therefore, it's important to support concurrent tracking performance in human-machine systems.

In addition to driving, the concurrent tracking performance is prevalent in many other scenarios, such as process control, in which an operator needs to control the flow of multiple material streams to balance process output; medicine, in which an anesthesiologist needs to adjust drug delivery pumps to control multiple physiological parameters; aviation, in which a pilot needs to track the speed and altitude of an airplane.

VII.2 Redundancy Gain and Cost

According to Multiple Resource Theory (MRT; Wickens, 2002), encoding information into redundant parallel modalities, such as visual, auditory, and tactile modalities, can mitigate mental resource competition in multitask scenarios and benefit multitask performance. For example, anesthesiologists who performed visual tasks while monitoring patient data showed better performance when patient data (e.g., blood pressure levels) were transformed into the auditory (Seagull, Wickens, & Loeb, 2001; Watson & Sanderson, 2004) or tactile (Ferris & Sarter, 2011) signals. The benefits of redundantly encoding information are called *redundancy gain* (Wickens et al., 2011).

In addition to redundant modality, encoding task-related information into redundant display dimensions can be another way to support multitask performance because it can disambiguate signals and increase adaptability in dimensional attention (*redundancy gain*). For example, Ardoin & Ferris (2016) found the tactile displays engaged redundant dimensions better supported dual task performance than the unidimensional displays that imposed larger competition for mental resources. However, encoding information into a redundant dimension does increase the complexity of signals and may interfere with the information processing and degrade task performance (*redundancy cost*; Wickens et al., 2011). In the same study, Ardoin & Ferris (2016) also observed that the tactile displays with redundant dimensions worsen the task performance compared to the unidimensional displays which had less competition with visual tasks.

VII.3 Evaluate the Effects of Novel Speedometer Displays on Concurrent Tracking Performance – Experiment 1

The study in this chapter aims to use the novel speedometer displays built in Chapter VI to produce redundancy gain and reduce redundancy cost in concurrent tracking task. Driven by this goal, this section answered the last research question (Q5) of the dissertation: **how do novel speedometer displays affect concurrent tracking performance in driving**. This question can be interpreted as two sub-questions:

1) **how can we improve concurrent tracking performance using novel speedometer displays that engage redundant modalities**;

2) **whether encoding information into redundant 'beat pattern' dimension of displays can further enhance concurrent tracking performance**.

The two questions were answered in Experiment 1 and 2, respectively. In Experiment 1, the novel speedometer displays included an "ambient-visual", an auditory, and a tactile speedometer display; in Experiment 2, the novel speedometer displays were the auditory and tactile displays that encode speed into the spatial location and beat pattern. Every display provided two pieces of information: 1) whether current speed is higher or lower than target speed (*speed direction*) and 2) the error between current speed and target speed (*speed error*). Both experiments were built on concurrent speed tracking and lane tracking task set in a simulated driving environment.

VII.3.1 Method

Participants

Twelve adults (9 Male, 3 Female; mean age 26 years old) from Texas A&M University participated in this experiment. All had normal or corrected-to-normal visual and auditory acuity, no known conditions that affected the tactile sensitivity of the back, and a valid driver's license for at least one year. The average driving experience was 3.7 years.

Apparatus

Four driving scenarios were created in the STISIM Drive® Driving Simulator. Participants interacted with the simulator via a Logitech G27 racing wheel and associated throttle and brake pedals. All scenarios used a standard vehicle cockpit view with an analog speedometer centered at the bottom of the screen (see Figure 30). The simulated roadway contained a set of consecutive curves of varying degrees that participants encountered in a pseudo-randomized order, so that each scenario was equivalent but unique. The curves were bordered by changing foliage and buildings. No other cars or obstacles were on the roadway. Throughout the scenario, a complex sinusoid-generated wind function increased or decreased the vehicle's speed within approximately 5 MPH. Each scenario was approximately 5 minutes long.

Figure 30. Driving simulation scenario, with the ambient-visual display projected to the background screen behind the monitor.

Each participant completed all four experimental conditions in a semi-counterbalanced order: (1) a "baseline" condition which did not include any novel displays and (2-4) three conditions which used one of the three displays (i.e., an "ambient-visual" display, an auditory display, and a tactile display. The speed was divided into nine ranges: one "on target" range (49-51 MPH) and eight ranges corresponding to four different distances from target speed in each direction (see Table 22). The presentation in each display was updated approximately 8 times per second.

The ambient-visual display was projected on a drop-down screen behind the computer monitor (see Figure 31). Warmer hues (e.g., red) indicated that the participant was driving above the target speed, and cooler hues (e.g., blue) indicated that the participant was driving below the target speed. Yellow was used as the "on target" hue because it was the midpoint on the visible spectrum between the warmer and cooler colors.

Table 22: Mapping Speed to the Three Novel Displays

	Speed (MPH)	ambient-visual (RGB codes)	Auditory (Hz)	Tactile (Location ; Hz)
↑	54>	(153, 0, 0)	690+702 (12)	7, 8; 250+262 (12)
	53 – 54	(255, 0, 0)	587+595 (8)	7, 8; 250+258 (8)
	52 – 53	(255, 102, 0)	493+497 (4)	7, 8; 250+254 (4)
	51 – 52	(255, 204, 0)	392+393 (1)	5, 6; 250+252 (1)
Acc	***49 – 51***	***(255, 255, 102)***	***349***	***3, 4, 5, 6; 250***
	48 – 49	(187, 239, 57)	329+328 (1)	3, 4; 250+249 (1)
↓	47 – 48	(0, 160, 0)	261+257 (4)	1, 2; 250+254 (4)
	46 – 47	(0, 102, 255)	220+212 (8)	1, 2; 250+258 (8)
	>46	(0, 0, 255)	174+162 (12)	1, 2; 250+238 (12)

Note. The target speed range is 49 -51 MPH. Acc represents "acceptable speed range". The RGB colors within the acceptable range were perceptually similar and clearly distinguishable from those outside of the acceptable range. Each number in the parenthesis after the pair of frequencies equals the difference between frequencies, indicating the beat frequency.

The auditory display consisted of a pair of ceiling speakers that displayed nine auditory presentations. The “on target” auditory presentation was a single pure tone at 349 Hz (approximately the pitch F4). Higher pitches (those above 349 Hz) indicated that the participant was driving above the target speed, and lower pitches (those below 349 Hz) indicated that the participant was driving below the target speed. Each of these “off target” pitch presentations was combined with another tone that of a slightly different pitch, generating an emergent phenomenon: binaural beats. For example, when driving at 52.3 mph, the auditory signal would be a combined 493 Hz + 497 Hz tone, resulting in a binaural beat. The frequency of beat is 4 Hz. The frequency of binaural beats increased discretely from 1 to 12 Hz as the speed deviated further from the target (See Table 22).

Tactile display consisted of eight C-2 tactors developed by Engineering Acoustics, Inc. All eight tactors were affixed to the inside of the back of a compression shirt and were spaced on the whole back to allow participants to distinguish easily the locations associated with “fast” (tactors 7 and 8 above the shoulder blades), on-target (tactors 3, 4, 5, and 6), or “slow” (tactors 1 and 2 near the iliac crests of the pelvis) (see Figure 31). An elastic weightlifting belt was used to enhance the skin contact with tactors 1-6. Tactors operated in a base frequency of 250 Hz and were combined with another frequency to generate “haptic beats”, a similar pattern of binaural beats, to indicate speeds outside of the target range. For example, when driving at 47.7 MPH, tactors 1 and 2 would be activated with a combined 250 + 246 Hz signal, forming a haptic beats at the frequency of 4 Hz. The frequency of haptic beats also went up when the speed further moved away from the target. The tactors vibrated at their maximum gain in order to ensure the perception of vibrations. In addition, noise-cancelling headphones were used in the tactile display condition to mute the audible sounds generated by the activating tactors.

Figure 31. Spatial locations of tactors on the inside back of a compression shirt.

Concurrent Tracking Task

There were two equally important tasks: lane tracking and speed tracking. The lane tracking task required participants to drive as closely as possible to the dashed lane marker. A piece of foam tape was placed directly on the desktop monitor to serve as the tracking "cursor" and indicate the center of the car. This "cursor" should pass directly over the lane marker. The speed tracking task required participants to keep the speed as close as possible to 50 MPH (target speed) by manipulating the throttle and brake pedals. Maintaining this target speed required offsetting the effect of wind and the accelerating and decelerating effects caused by hills in the driving scenario. The engine sounds were muted so that they could not be used as speed cues.

Procedure

After signing the consent form and completing a background questionnaire, participants got familiar with each type of display through our brief instructions and demonstrations of displayed presentations and how they mapped to the speed information. They then practiced the concurrent tracking task under each display condition in a 15-minute training session. After training, participants completed four experimental (one baseline condition and three display conditions). There was a 3-minute interval between consecutive experimental conditions but additional breaks were also allowed. In the end, participants filled a feedback questionnaire for evaluating the displays and ranked them in terms of the overall preference.

Analysis of Results

Performance was measured in two primary dependent variables: lane deviation and speed deviation. Lane deviation was the mean difference in feet from the center of the vehicle to the target lane marker over a scenario. Similarly, speed deviation was the mean difference in MPH from actual speed to the target speed over a scenario. Each scenario began with a 400-foot “ramp-up” section that was a stretch of straight road and ended at a 50 MPH speed limit sign. It allowed participants to accelerate to the target speed, therefore the data in this section was not considered in the analysis.

As an additional dependent measure, the “acceptable performance percentage” (AP%) was derived from the two primary dependent variables. It was calculated as the percentage of time that the participant drove within the acceptable speed range (48 - 52 MPH) while also maintaining the center of vehicle (indicated by foam tape) within +/- 0.5 feet of the center of the dashed line. This range was chosen because the dashed line was 1.0 foot wide. With approximately the same width, the foam tape “cursor” that represented the center of the vehicle should overlap the dashed line as much as possible.

The subjective ratings of each display were generated according to five attributes: 1) satisfaction, which was the extent to which displays supported speed tracking task; 2) reliance, which represented the extent to which participants relied on the displays; 3) interpretation, which denoted the mental efforts used to interpret the displays; 4) distraction, which was the extent of distraction on lane tracking caused by displays; 5) annoyance, which indicated how annoying participants felt when they process the displays. For each attribute, a rating of “1” represented the worst case (not satisfied; didn’t rely on display at all; extremely difficult to interpret; extremely distracting; extremely annoying), while a rating of “10” represented the best case.

In data analysis, one-way repeated measures ANOVAs were conducted in R to determine whether the experimental conditions significantly affected the dependent variables. Tukey's Honest Significant Difference (HSD) tests were performed to determine the differences in dependent variables among experimental conditions. In the end, subjective ratings of displays were analyzed using one-way repeated measures ANOVAs, and preference rankings of displays were also analyzed using a Friedman's test.

VII.3.2 Results

Speed Deviation

The speed deviation was significantly affected by experimental conditions ($F(3, 33)=11.71$, $p<0.001$, $\eta^2=0.239$). Post-hoc tests showed that speed deviation was significantly higher in the baseline condition (3.16 MPH) than in the ambient-visual (2.45 MPH, $p=.001$), Auditory (2.10 MPH, $p<.001$), and Tactile (2.49 MPH, $p=.002$) conditions (see Figure 32).

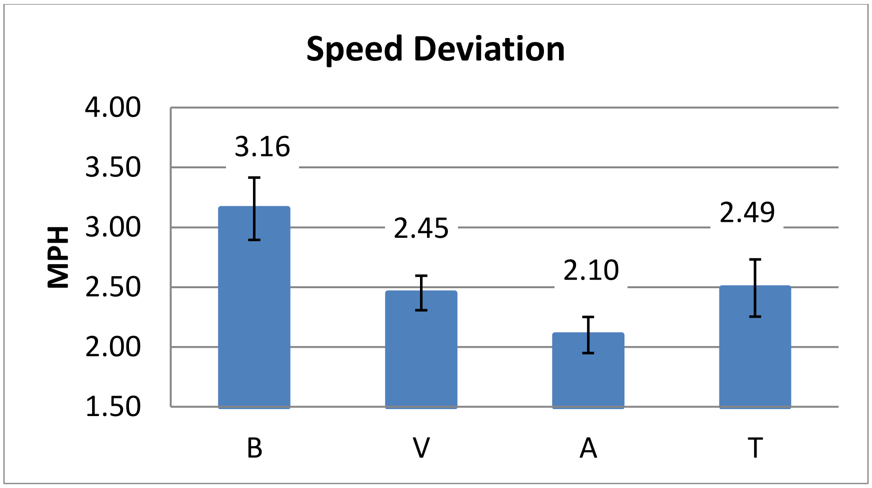


Figure 32. Speed deviation (in MPH) for each experimental condition. B, V, A, and T represent baseline, ambient-visual, auditory, and tactile, respectively. Error bars represent standard error.

Lane Deviation

The lane deviation wasn't significantly affected by experimental conditions (F (3, 33)=1.63, *p*=0.202, η^2=0.043) (see Figure 33).

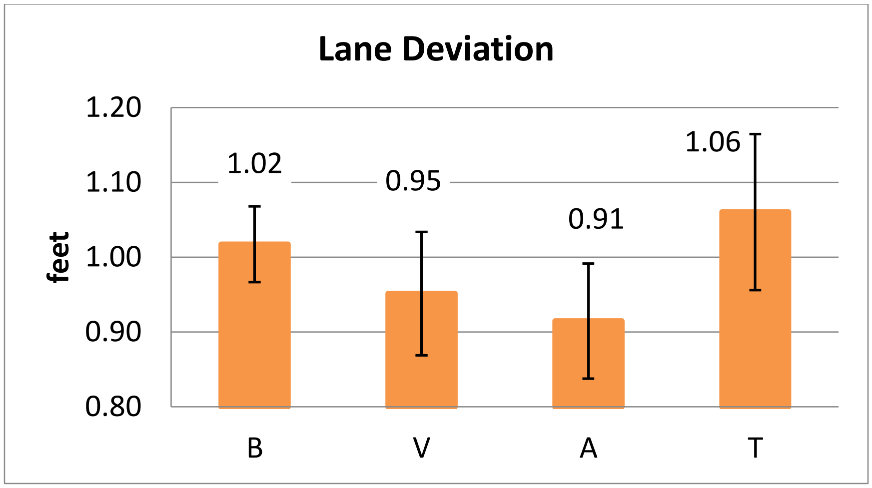


Figure 33. Lane deviation (in feet) for each experimental condition. B, V, A, and T represent baseline, ambient-visual, auditory, and tactile, respectively Error bars represent standard error.

Acceptable Performance % (AP%)

The AP% was significantly impacted by experimental conditions (F(3, 33)=7.61, p=.001, η^2=0.140). Post-hoc tests showed that AP% for the baseline condition (20%) was significantly worse than that in the ambient-visual condition (26.6%, p=.002) and the Auditory condition (28.2%, p<.001) displays. AP% in the Auditory condition was also significantly higher than in the Tactile condition (22.9%, p=.027) (see Figure 34).

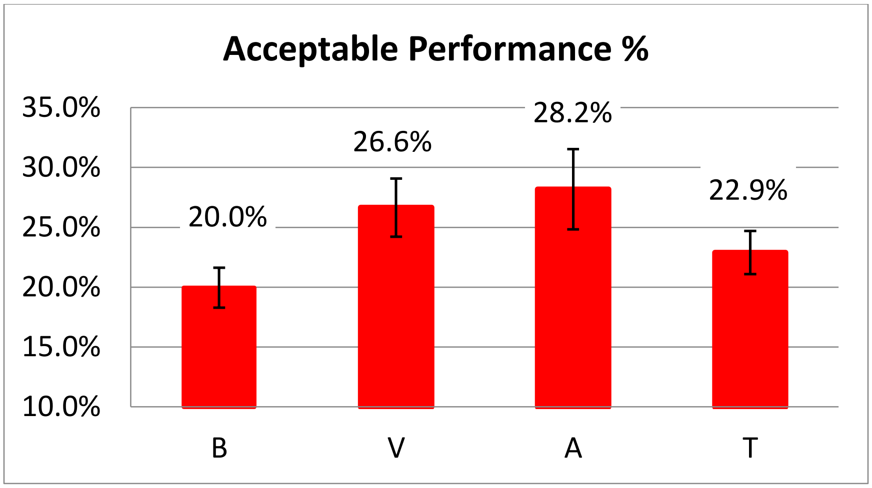


Figure 34. The Acceptable Performance % for each experimental condition. B, V, A, and T represent baseline, ambient-visual, auditory, and tactile, respectively. Error bars represent standard error.

Subjective Feedback

The average ratings for each display are given in Table 23. The effect of experimental condition was significant on the ratings of satisfaction (F(3,33)=13.77, p<0.001, η^2=0.398), distraction (F(3,33)=9.373, p<0.001, η^2=0.378), and annoyance (F(3,33)=3.545, p=0.025, η^2=0.140). For both Satisfaction and Distraction, post hoc test showed that baseline display had lower ratings than other displays (p for Satisfaction: ambient-visual=.001, Auditory<.001, Tactile<.001; p for Distraction: ambient-visual=.002, Auditory<.001, Tactile<.001). For annoyance, the baseline display showed higher rating score than the Auditory (p=.025) and Tactile (p=.024) displays.

Table 23: Average Ratings for each Type of Display

Attribute	Ratings	Significance
Satisfaction	A: 8.00; T: 7.50; V: 6.83; B: 4.92 (A, T, V > B)	< .001
Reliance	A; 7.75; T: 7.58; V: 6.67	Not Significant
Interpretation	A: 6.50; T: 6.33; V: 6.08; B: 5.33	Not Significant
Distraction	A: 7.50; T: 6.75; V: 6.50; B: 3.67 (A, T, V > B)	< .001
Annoyance	B: 8.00; V: 6.42; A: 5.67; T: 5.67 (B > A, T)	.025

Note. B, V, A, and T represent Baseline, ambient-visual, Auditory, and Tactile, respectively. The results of post hoc tests are showed inside the parentheses.

The average rankings of overall preference to the experimental conditions were significantly different ($\chi^2(3, 12)=14.1$, $p=.003$) (see Figure 35). The auditory condition was ranked as the best followed closely by the tactile condition. The baseline condition was ranked as the worst.

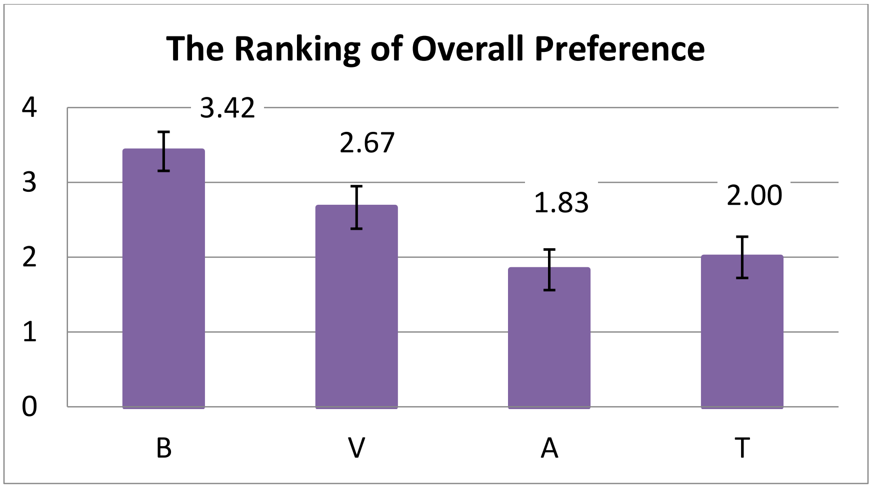


Figure 35. The ranking of overall preference for each experimental condition (lower score represent higher preference). B, V, A, and T represent baseline, ambient-visual, auditory, and tactile, respectively. Error bars represent standard errors.

VII.3.3 Discussion

Since distributing information to parallel sensory channels can support human information processing, in Experiment 1 we applied the redundant ambient-visual, auditory, and tactile speedometer displays to convey speed information to drivers (at the same time with the baseline display). The study illustrated the effects of these redundant speedometer displays on concurrent lane-and-speed tracking performance and revealed participants' subjective feedback of using these displays.

The results demonstrated that all novel speedometer displays, compared to the baseline display, significantly improved the speed tracking performance. Among them the ambient-visual and auditory displays generated the best performance. But none of them significantly affect the lane tracking performance, which suggests that the improvements in speed tracking performance did not appear at the cost of degrading the lane tracking performance. Moreover, the ambient-visual and auditory displays (not including tactile display) also significantly improved the overall tracking performance according to the results of AP%. The findings were consistent with the subjective feedback in which participants showed their stronger preference (which was based on higher rating on satisfaction sacle and lower rating on distraction scale) to the novel speedometer displays than the baseline speedometer display alone in the concurrent tracking task.

Although not significant, the auditory display is the best one to support concurrent tracking performance, leading to the lowest speed deviation and the highest AP%. According to Multiple Resource Theory (MRT; e.g., Wickens, 2002), auditory modality is parallel to visual modality so that auditory modality can share the visual load without interfere with visual modality. Therefore, using the auditory display to present redundant (speed) information can reduce the original competition for limited visual resources and improving concurrent tracking performance.

The ambient-visual display is the second best display to support the speed tracking and overall performance. Similar to the auditory modality, peripheral vision engaged by the ambient-visual displays is also considered to be parallel to focal vision (engaged by the baseline speedometer display), according to MRT (Wickens, 2002). However, the ambient light emanating from the background screen was bright and changing, generating the tendency to reorient participants' focal vision away from the monitor. This phenomenon of redirecting visual attention is called *phototropism* (Wickens, Lee, Liu, &

Gordon-Becker, 2004). Because of the effect of phototropism, the ambient-visual display didn't improve the task performance as much as the auditory display.

The tactile display benefited only the speed tracking performance yet not the overall performance. According to MRT, tactile modality is also parallel to the visual modality but it may be not as capable as auditory and visual modalities in terms of information processing capacity. First, human only have limited bandwidth of tactile perception compared to visual and auditory perception. Second, participants may be still unfamiliar with the tactile devices because the tactile devices haven't been broadly used in vehicles. On the opposite, the visual and auditory displays can be found in almost every car.

The tactile display may also hinder the information processing at the cognitive stage. Processing the tactile tracking information required the use of spatial working memory to determine *where* the vibrations originated on the body (similar to *where* the speedometer indicator was relative to the target 50 MPH). Meanwhile, the lane tracking task also demanded spatial working memory, causing the competition for spatial working memory at the cognitive stage. This competition may offset the tactile display's effort to reduce the attentional conflicts at the perception stage, degrading the concurrent tracking performance. Unlike the tactile display, the ambient-visual and auditory displays primarily required *symbolic* processing at the cognitive stage, with no disruption on the spatial processing that was engaged by the lane tracking task. Therefore, it's important to consider sensory modalities and encoding dimensions that cause less cognitive interference for display design (Ferris & Sarter, 2010).

In addition to the benefits of novel speedometer displays, we also need to notice the fact that these displays were much more annoying than the baseline display according the subjective feedback. This annoyance of display is a critical usability issue that has to be solved in the following study.

VII.4 Evaluate the Effects of Novel Speedometer Displays on Concurrent Tracking Performance – Experiment 2

VII.4.1 Method

Participants

The study recruited 15 adults (9 Male, 6 Female; mean age 24 years old) from Texas A&M University. All participants had a valid driver's license for at least one year. They also had normal or corrected-to-normal visual and auditory acuity and no known conditions affecting the tactile sensitivity of the back. The average driving experience was 3.3 years.

Apparatus

The driving scenarios and the concurrent tracking task used the same setup as in the Experiment 1. The Experiment 2 examined the supporting effect of the multi-dimensional display in a higher workload situation. Hence, we increased the cognitive demand for the lane-tracking task by narrowing down the acceptable "speed range" from previous 48 -52 MPH to 49 - 51 MPH.

The multi-dimensional displays were modified from the Auditory and Tactile displays developed in the Experiment 1. The new auditory display engaged spatial dimension and the new tactile display used a different spatial layout on the back. On the basis of using spatial dimension, beat pattern was also employed into auditory and tactile displays, generating two displays that engaged redundant 'spatial+beat' dimensions. Therefore, there were four types of displays: auditory-spatial, auditory-redundant, tactile-spatial, and tactile-redundant. The presentations of each display naturally mapped to nine speed levels (see Table 24).

Table 24: The Nine Speed Levels and Their Respective Auditory and Tactile Presentations

Speed (MPH)	A-spatial (loca; Hz)	A-redundant (loca; Hz)	T-spatial (loca; Hz)	T-redundant (loca; Hz)
54>	1; 690	1; 690+702(12)	8 ; 262	8 ; 250+262(12)
53 - 54	2; 587	2; 587+595(8)	7, 8; 258	7, 8; 250+258(8)
52 - 53	3; 493	3; 493+497(4)	6, 7; 254	6, 7; 250+254(4)
51 - 52	4; 392	4; 392+393(1)	5, 6; 251	5, 6; 250+251(1)

Table 24: Continued

	Speed (MPH)	A-spatial (loca; Hz)	A-redundant (loca; Hz)	T-spatial (loca; Hz)	T-redundant (loca; Hz)
Acc	**49 - 51**	**5; 349**	**5; 349**	**4, 5; 250**	**4, 5; 250**
	48 - 49	6; 329	6; 329+328(1)	3, 4; 249	3, 4; 250+249(1)
	47 - 48	7; 261	7; 261+257(4)	2, 3; 246	2, 3; 250246(4)
	46 - 47	8; 220	8; 220+212(8)	1, 2; 242	1, 2; 250+242(8)
	>46	9; 174	9; 174+162(12)	1 ; 238	1 ; 250+238(12)

Note. 'Acc' represents acceptable speed range. 'loca' represents location and 'Hz' the pitch or activation frequencies. 'A' represents auditory and 'T' represents tactile. Each number in the parenthesis after the pair of frequencies equals the difference between frequencies, indicating the beat frequency.

The auditory-spatial display used a pair of noise-cancelling headphones playing sounds from nine virtual origins. The perception of virtual origins was created by SLAB (a software developed by NASA Ames Research Center) which manipulated the inter-ear intensities (Wolfe, Dluender, & Levi, 2011). The nine virtual origins were equally distributed as if originating from a semicircle on the transverse plane at the height of ear (see Figure 36). The radius of the semicircle was 50 cm. Since human only has limited ability to discriminate auditory spatial origins, the pitch of each auditory presentation was adjusted to increase from location 1 to 9 (see Table 24) to enhance the auditory spatial perception.

The auditory-redundant display was modified from the auditory-spatial display by involving an additional beat pattern. The beat pattern was generated by combining another slightly different pitch into the auditory display. When speeds moved away from the target towards either extreme, the frequency of beats discretely raised from 1 Hz to 12 Hz (see Table 24). The “on target” speed level was still denoted by a pure tone. In addition, the volumes of both auditory displays were tuned at a level that was comfortable for participants and can be reliably heard.

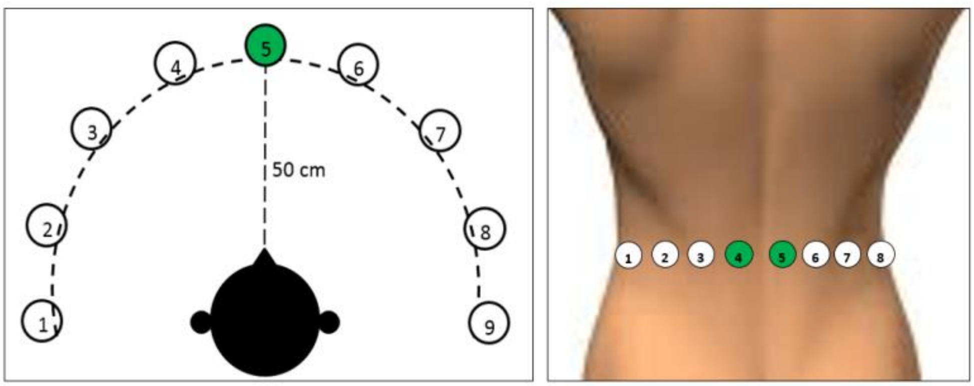


Figure 36. The nine virtual auditory locations surrounding the head (left) and the eight tactor locations on the lower back (right). The activation on greened locations indicates the ongoing target speed.

The tactile-spatial display consisted of eight C-2 tactors which were arranged horizontally and symmetrically across the lower back. The tactors were doubly secured with an elastic weightlifting belt and a strap on the top of the belt. We used the maximum gain of tactile vibration to ensure that participants felt the vibrations. Similar to the auditory display, the tactile vibrations were displayed continuously to convey the current speed.

The tactile-redundant display added haptic beats into the tactile-spatial presentation. Like the auditory-redundant display, the frequency of haptic beats increased from 1 Hz to 12 Hz as the site of vibrotactile presentation moving away from the center to either side of the back.

Procedure

After signing the consent form and completing the background questionnaire, participants were then familiarized with all four experimental displays in a training session. In the training, participants were required to identify the virtual locations of randomly-presented sounds until they achieved their best accuracy. Participants then completed four training scenarios with one under each display condition, which took a total of 15 minutes. After the training, each participant completed five experimental conditions in a semi-counterbalanced order, including one Baseline and four display conditions. Each condition lasted 5 minutes with a 2-minute break after it. At the end of the study, participants completed a feedback questionnaire to rate and rank the Baseline and experimental displays.

Analysis of Results

Driving performance was assessed using the same dependent variables as in the Experiment 1: lane deviation, speed deviation, and acceptable performance percentage (AP%). As stated, the speed range for AP% was modified from the Experiment 1 to be 49~51 MPH. The three dependent variables and the subjective ratings were analyzed in one-way repeated measures ANOVAs in R. Tukey's HSD were used to determine differences in the dependent variables among experimental conditions. Friedman's test was used to analyze the overall rankings of displays.

VII.4.2 Results

Speed Deviation of Baseline Condition

The speed deviation of baseline condition in the Experiment 1 was significantly higher than that in the Experiment 2 ($F(1,25)=4.80$, $p=.038$, $\eta^2=0.161$). See Figure 37.

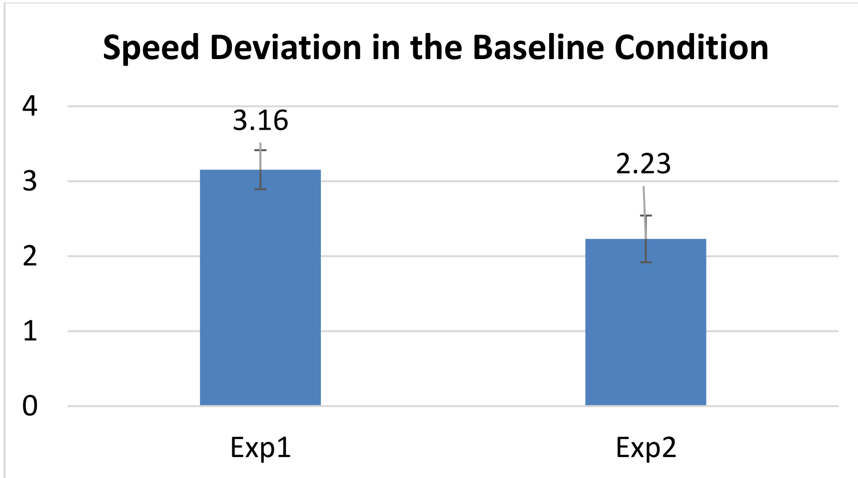


Figure 37. The speed deviation of baseline condition in both experiments.

Speed Deviation

Experimental condition significantly affected speed deviation ($F(4, 56) = 5.54$, $p<.001$, $\eta^2=0.168$). The speed deviation in baseline condition (2.23 MPH) was significantly higher than those in the display conditions (auditory-spatial: 1.54 MPH, $p=.029$; auditory-redundant: 1.18 MPH, $p=.001$; tactile-spatial: 1.42 MPH, $p=.006$; tactile-redundant: 1.42 MPH, $p=.006$). See Figure 38.

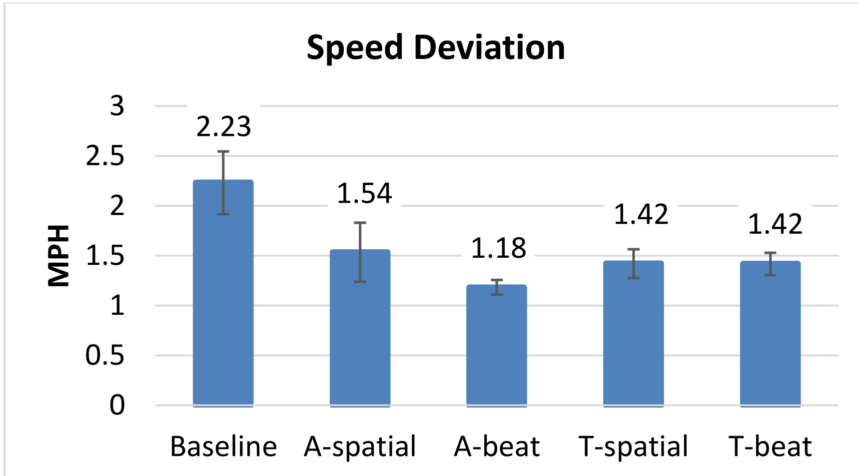


Figure 38. Speed deviation (in MPH) across all experimental conditions. A represents auditory and T represents tactile. Error bars indicate standard error.

Lane Deviation

Lane deviation did not differ significantly across experimental conditions ($F(4, 56) = 0.251$, p=.908, η^2=0.002). See Figure 39.

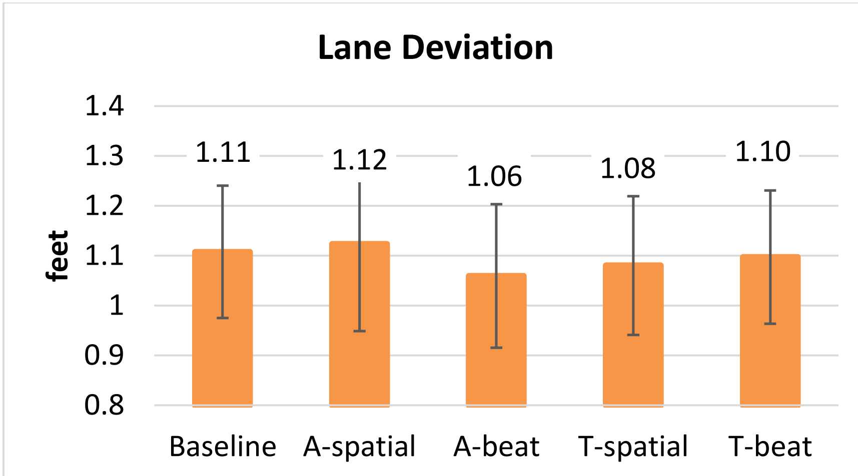


Figure 39. Lane deviation (in feet) across all experimental conditions. A represents auditory and T represents tactile. Error bars indicate standard error.

Acceptable Performance % (AP%)

The AP% varied significantly across experimental conditions ($F(4, 56) = 7.09$, $p<.001$, $\eta^2=0.083$). Post hoc tests showed that AP% in the baseline condition (AP%=23.0%) was significantly lower than in auditory-spatial (AP%=30.8%, $p=.001$), auditory-redundant (AP%=32.8%, $p<.001$), and tactile-spatial (AP%=29.0%, $p=.021$) conditions. See Figure 40.

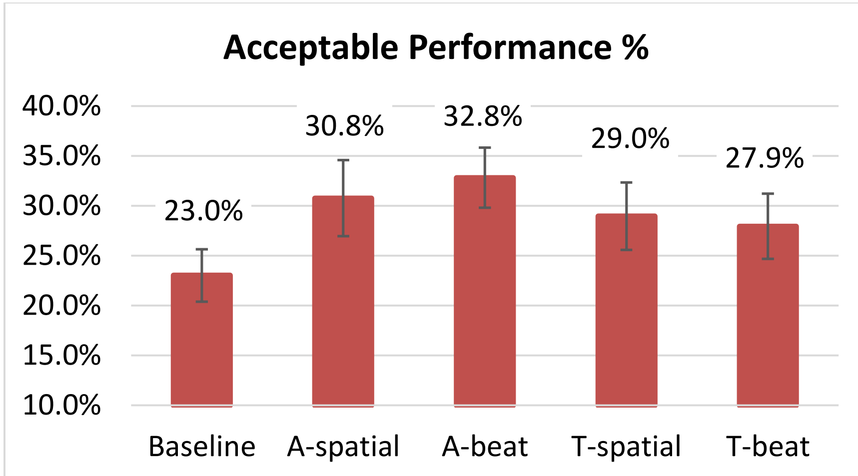


Figure 40. Acceptable Performance % across all experimental conditions. A represents auditory and T represents tactile. Error bars indicate standard error.

Subjective Feedback

Significant differences were found among displays for the ratings of satisfaction ($F(4, 56) = 2.90$, p=.030, η^2=0.115), distraction ($F(4, 56) = 5.028$, p=.002, η^2=0.215), and annoyance ($F(4, 56) = 4.02$, p=.006, η^2=0.145) (see Table 25). Post hoc tests showed that the satisfaction rating for the auditory-redundant condition was significantly higher than for the Baseline (p=.017) condition. The distraction rating for the baseline condition was significantly lower than for all display conditions (auditory-spatial: p=.036; auditory-redundant: p=.011; tactile-spatial: p<*.001*; tactile-redundant: p=.005). The annoyance rating of baseline condition was significantly higher than for all display conditions (auditory-spatial: p=.017; auditory-redundant: p=.010; tactile-redundant: p=.006). In the end, overall preference rankings for displays did not significantly differ from each other in a Friedman's test.

Table 25: The Average Ratings for each Display

Attribute	Ratings	Significance
Satisfaction	A-redundant: 8.00 ; T-spatial: 7.73 ; T-redundant: 7.27; A-spatial: 6.93; B: 5.87 (A-redundant > B)	.030
Reliance	A-redundant:7.93; T-spatial: 7.93; T-redundant: 7.86; A-spatial: 7.00	n.s.
Interpretation	A-redundant: 6.67; T-redundant: 6.27; T-spatial: 6.20; B: 5.86; A-spatial: 5.60	n.s.
Distraction	T-spatial: 7.33; T-redundant: 6.80; A-redundant: 6.60; A-spatial: 6.27; B: 3.80 (A-spatial, A-redundant, T-redundant, T-redundant > B)	.002
Annoyance	B:7.89; T-spatial: 6.00; A-spatial: 5.47; A-redundant: 5.33; T-redundant: 5.20 (B > A-spatial, A-redundant, T-redundant)	.006

Note. A, T and B represent Auditory, Tactile and Baseline respectively. The results of post hoc tests are listed inside the parentheses at the bottom.

VII.4.3 Discussion

Experiment 2 investigated whether the more disambiguating yet more complex display can further improve driver performance. This was addressed by examining how multi-dimensional displays affected concurrent tracking task performance. The information for speed tracking was encoded into the spatial dimension or redundant 'spatial+beat' dimension of the auditory and tactile displays. The findings of Experiment 2 improve our understanding on how different display dimensions affect driver's concurrent tracking performance.

The lane deviation in Experiment 2 was significantly smaller than that in Experiment 1 and it was not different among experimental conditions, which suggested that the participants may attribute more mental resources to the lane-tracking tasks in Experiment 2. Comparing to the lane-tracking performance, the speed-tracking performance was significantly improved by all four novel speedometer displays. In addition to performance, the questionnaire feedback showed that the novel speedometer displays caused less distraction on the lane tracking task. These objective and subjective findings again demonstrated that the novel speedometer displays are able to reduce the competition for focal-visual resources and support the concurrent tracking performance.

Although human only has limited auditory spatial perceptual capacity, the auditory-spatial display still supported better tracking performance than the baseline display. Perhaps, pitch, which was also engaged into the auditory-spatial display, severed as a salient and robust auditory feature to enhance the perception of auditory tones. Thus, symbolic encoding associated with pitches could occur at the same time with the spatial encoding. This combination of symbolic encoding and spatial encoding in a single display modality can generate redundancy gain in information perception (Ardoin & Ferris, 2016).

Comparing to the auditory-spatial display, the auditory-redundant display did not further improve the speed tracking performance. But participants were more satisfied with the auditory-redundant display than the auditory-spatial display. This difference between objective performance and subjective feeling suggested that the auditory-redundant display may benefit the cognitive process but not the manual performance. Moreover, the auditory-redundant display may only produce perceptual benefits for those who were difficult to discriminate the pure tones in the auditory-spatial display. For those who were good at recognizing pitch, they may not obtain additional benefits from the auditory-redundant display.

Tactile-spatial display also lead to better concurrent tracking performance than the baseline display. The vibrations on the lower back enabled the participants to monitor the speed information intuitively. Since the spine serves as an anatomical reference point (Cholewiak, Brill, & Schwab, 2004), tactors 4 and 5 near the spine were the prominent spots to indicate the on-target status, and the smooth left-and-right movements of vibration along the 'line of tactors' naturally mapped to the moving index in the Baseline speedometer. However, according to participant feedback, the user's sensitivity to the vibrations at tactors 4 and 5 decreased over time, potentially illustrating a neural inhibition effect that occurs when sensory neurons become fatigued under repeated vibrations. Practically, this reported insensitivity to tactile vibration actually could serve as an "anti-signal" to the time when driving at the target speed level. Thus, the sensitivity of vibrotactile locations and the neural inhibition effects may contribute together to the effectiveness of the tactile-spatial display.

As with the auditory-spatial versus auditory-redundant displays, the tactile-redundant display also failed to improve speed tracking performance over the simpler tactile-spatial display. Moreover, the tactile-redundant display did not significantly improve the AP% over the baseline display. It was observed that the driving performance of two participants were severely interrupted by the tactile-redundant display, especially when the vibration traveled to the both sides of the back. This may be another way in which redundancy cost was expressed, essentially overriding any redundancy gain that may be observed with the tactile spatial+beat pattern.

VII.5 General Discussion

Chapter VI answered the question of how do the novel speedometer display affect continuous concurrent tracking task (Q5) since concurrent tracking task become more and more important in many fast-paced and data-rich work environments.

There modalities - ambient-visual, auditory, and tactile - were individually engaged by the novel speedometer displays. These modalities are parallel to focal vision according to the Multiple Resource Theory (Wickens, 2002). Hence, the novel speedometer displays caused less interruption on the lane tracking performance (which required focal vision), reduced visual resource competition, and better support concurrent tracking performance. However, it’s interesting to notice that although the auditory speedometer displays showed the highest cognitive efficiency in Chapter VI, it did not lead to the best tracking performance in the study of this chapter.

In addition to modality, this chapter also explored the effects of display dimension on concurrent tracking performance. It was found that the auditory and spatial displays effectively improved the tracking performance by encoding speed information into their spatial dimensions. However, redundantly encoding speed into auditory beats or haptic beats did not further enhance the concurrent tracking performance in current experiments. But the use of redundant beat dimension improved the cognitive efficiency of the auditory display (See Chapter VI).

The interesting finding that the auditory-redundant display with higher cognitive efficiency didn't contribute to better (concurrent) tracking performance conflicts with what we found in Chapter IV, which showed that the more efficient displays associated with better task performance. Perhaps participants can obtain 'extra' information from the baseline displays when it's difficult to perceive the information from the auditory-spatial display, resulting in concurrent tracking performance as good as that in the auditory-redundant display condition. Therefore, there is a need to exclude the 'hidden' benefits of the baseline display in the study that removes this display.

VII.6 Conclusion

Chapter VII applied the novel speedometer displays that were evaluated in the previous chapter to support concurrent tracking performance in driving. It demonstrated that the novel speedometer displays which encoded speed information into redundant modalities or dimensions enhanced the concurrent lane-and-speed tracking performance in driving simulation scenarios. The results of this chapter provided insight into the gain and cost of adopting each redundant modality and dimension, which are valuable for the design of display systems in a wide range of multitask scenarios. Moreover, this chapter found that the novel speedometer displays that showed higher cognitive efficiency (in the previous chapter) did not always enhance the concurrent task performance as much as we expected, which illustrated the complexity in the relationship between cognitive efficiency and multitask performance.

CHAPTER VIII

SUMMARY AND CONCLUSION

The dissertation integrated my eight research projects over the past five years (2011-2016) to illustrate *cognitive efficiency,* a critical property of display to address information overload challenge in human-machine systems. The exploration of this property engaged the knowledge of several important human factors topics, including sensation/perception, mental workload measure, display design, and multitask performance. In my dissertation, cognitive efficiency was primarily investigated from three perspectives: its measurement techniques, its relationship with multitask performance, and its application on novel speedometer display in driving.

VIII.1 Cognitive Efficiency Metric

As a measurement technique, Cognitive Efficiency (CE) metric is the core of this dissertation. It was proposed for quantitatively evaluating cognitive efficiency of display. CE metric consists of two dimensions: *display informativeness* and *required mental resources*. From a systematic view, the measure of each dimension is defined as a function that depends on characteristics of the human (h), display (d), and task/environmental context (k). Built upon the measure of each dimension, CE metric is formulated as the ratio of display informativeness to required mental resources (See the equation below).

$$\text{CE}(h_i, d_i, k_i) = \frac{display\ informativeness\ (h_i, d_i, k_i)}{required\ mental\ resources(h_i, d_i, k_i)}$$

Different from previous measures of display efficiency, the proposed CE metric provides a comprehensive interpretation of display efficiency by taking into account all relevant factors in human-machine systems. For example, cognitive efficiency in context k_i can then be compared between displays (with the same human, h_i) or between humans (with the same display, d_i). Therefore, CE metric largely expands the existing interpretation of display efficiency in human-machine systems.

The measures of each dimension in CE metric were explored in Chapter II. In this chapter, display informativeness was indicated by information transmitted according to Information Theory. This indicator can reliably show the quantity of discrete information delivered from a display to an observer. However, it's difficult to calculate information transmitted in real-world scenarios because it's impractical to obtain the a priori condition probabilities (which is required for calculation) outside the constructed experimental environments (Xie & Salvendy, 2000). Moreover, the computation of information transmitted requires the feedback from human by verbal report or button pressing, which may interrupt other ongoing tasks. Thus, we need to find more practical and nonintrusive ways to evaluate this dimension of CE metric in later study.

In Chapter II, physiological measure was mainly used as the assessment of required mental resources in CE metric because physiological measure can provide continuous and high-resolution monitoring of cognitive states. In the study of Toyota

Economics Settlement Safety Research described in this chapter, we used HRV and SCR successfully detect the cognitive states under three types of short-term loads (i.e., mental, emotional, and motoric) and an acute stress event (i.e., unintended acceleration). The study showed physiological measure can be a promising way to measure the required mental resources in CE metric. In the future, the quality of physiological measure is expected to be promoted by the development of wearable sensors embedded in machines (e.g., heart rate monitor installed inside steering wheel to indicate driver's mental workload).

In addition to assessing mental workload under normal conditions, physiological measure may also be able to detect the occurrence of cognitive redline, which indicates mental overload (information overload). The last study in Chapter II showed that HRV stayed at its lower limit when/after human mental workload approached its cognitive redline. Therefore, the lower (or upper) limit of physiological measure may strongly associate with cognitive redline and be able to support the detection of mental overload. The finding may provide effective way to identify information overload syndrome but needs to be validated with further efforts.

Some issues of the measure of required mental resources are still unsolved in this dissertation. First, it's still not quite sure how to accurately measure the mental workload that is only imposed by a single subtask in multitask scenarios because it's difficult to divide overall mental workload into several portions that correspond to each subtask. Since the relationship between measured workload and actual workload can be nonlinear (Estes, 2015), a simple summation of all partial workload measures cannot accurately represent overall mental workload. Further efforts are needed to address this issue of partial workload measure.

Second, the sensitivities of mental workload measures largely depend on the nature of tasks and workload levels, which means the sensitivity may be unreliable under different task conditions. Perhaps, we can algorithmically combine multiple objective and subjective measures to improve the robustness of workload assessment, which will require the interdisciplinary efforts from various fields, such as human factors, computer science, and neuroscience.

CE metrics, which was built on the measures described in Chapter II, were initially validated with the basic displays in Chapter III. The results showed that cognitive efficiency of the visual displays was higher than that of the auditory displays in the large information content. Moreover, cognitive efficiency became greater when the display encoded information into its spectrum dimensions (e.g., hue and pitch) than intensity dimensions (e.g., brightness and loudness). However, it was interesting to find that the value of cognitive efficiency also depends on the measures selected for each dimension of CE metric. For example, CE metric that relied on subjective measures of required mental resources (e.g., NASA TLX) generated a different pattern of cognitive efficiency than that used objective measures (e.g., EEG). This initial investigation of CE metric encouraged my further investigation of different versions of CE metric under multitask conditions in Chapter IV.

VIII.2 Cognitive Efficiency and Multitask Performance

Chapter IV further explored the value of CE metric by investigating its relationship with multitask performance. This relationship is complicated because the two dimensions - display informativeness and required mental resources - of CE metric influence multitask performance in opposite ways. Higher display informativeness can improve task performance while larger imposed mental workload degrades task performance. Chapter IV compared several different CE metrics (which adopted different mental workload measures) in terms of their predictive power of multitask performance. The results demonstrated that all CE metrics positively correlates with multitask performance and the CE metric engaging subjective mental workload measures (NASA TLX) showed the strongest correlation. This finding provided an important theoretical support for the use of CE metric in multitask environments: the cognitively efficient display enhances multitask performance.

In addition, Chapter IV found that inter-individual human factors significantly influenced the sensitivity of CE metric and their prediction of multitask performance. CE metric was more predictive of multitask performance in the group with lower multitask performance (low-performing group). Since the low-performing group was found not as cognitively efficient as the high-performing group, it's possible that this group was not able to compensate for the low efficiency of a display, thus being more sensitive to display efficiency.

Future studies of interest derived from Chapter IV include:

- examining how the performance-predicting power of CE metrics varies between extreme high (at one's cognitive redline) and low workload levels;
- identifying more human and contextual factors that affect cognitive efficiency;
- applying CE metric to more complex real-world displays and tasks.

VIII.3 Cognitive Efficiency of Novel Speedometer Display

Based on the findings of previous chapters, the dissertation moved on to discussed how we can evaluate the cognitive efficiency of novel speedometer display. Before stepping into the study on novel speedometer display, we investigated a special display dimension: beat pattern. The study in Chapter V illustrated the neural basis behind the perception of haptic beats (the beat pattern of haptic modality). Moreover, the study found that the perceptual sensitivity of haptic beats was affected by sex, body location, and beat frequency, which provided valuable guidance for the use of beat pattern in the design of novel speedometer display, such as using high frequency of beats to indicate larger deviation between the current and target speeds.

A novel speedometer display was designed to convey continuously speed information to drivers via a modality that was parallel to focal vision, such as ambient-visual (engaging peripheral vision), auditory, or tactile modality. In Chapter VI, the novel speedometer displays were evaluated using CE metric which can quantify the informativeness of continuous display. The study found that the cognitive efficiency of the ambient-visual, auditory, and tactile displays was at the same level, but the ambient-visual display was less informative than the auditory and tactile displays, suggesting that it may require less mental resources. The different patterns between efficiency and informativeness measures, again, illustrate the need to take into account the measure of both display informativeness and required mental resources for a comprehensive evaluation of display.

The study in Chapter VI also examined whether we can enhance the cognitive efficiency of the auditory and tactile displays by encoding speed information into the redundant beat pattern. It's found that the use of beat pattern largely increased the cognitive efficiency of the auditory display, but not of the tactile display (because beat pattern enhanced the informativeness of the auditory display but not the tactile display). This finding showed that the cognitive efficiency of a display modality largely depends on which dimensions of the engaged modality are selected to encode information.

After analyzing the novel speedometer displays, in Chapter VII, we moved one step further to use these displays to support real-time concurrent tracking task which exits in many fast-paced and data-rich environments (e.g., driving, process control, and medicine). Implementing redundant novel speedometer displays in vehicle further enhanced the concurrent speed-and-lane tracking performance than using the traditional speedometer display along. However, the tracking performance was not significantly different under the ambient-visual, auditory, and tactile display conditions, which was consistent with their cognitive efficiency measure in Chapter VI. Moreover, engaging the beat pattern into the auditory or tactile presentation did not benefit the concurrent tracking performance, although it improved the cognitive efficiency of the auditory speedometer display in Chapter VI. In addition, participants showed stronger preference to use novel speedometer display for concurrent tracking task, but they also reported high annoyance of these displays.

Based on these findings, the study of novel speedometer display introduced several directions of future research, including:

- investigating the differences in the task performance between student groups of different fields (e.g., engineering vs. psychology);
- exploring other dimensions of each display and determining the best one to improve cognitive efficiency of display in human-machine systems;
- eliminating the cognitive interference caused by the novel speedometer displays;
- generalizing the use of novel speedometer displays to an ecological environment.

In conclusion, my dissertation has been an intellectual journey of understanding the quantitative evaluation of cognitive efficiency of display. It discussed how Cognitive Efficiency metric can be used to determine display characteristics that are less likely to produce instances where users are overwhelmed by information. Furthermore, it explored CE metric's predictive power of human operational performance in multitask scenarios, such as driving. In the end, CE metric was examined on novel speedometer displays which were used to support concurrent tracking performance. While the accuracy of cognitive efficiency measurements was still limited due to the quality of current measurement technologies, this dissertation offered deeper insight into the quantitative evaluation of display components in the human-machines systems.

REFERENCES

Aldrich, T. B., Szabo, S. M., & Bierbaum, C. R. (1989). The development and application of models to predict operator workload during system design. In *Applications of human performance models to system design* (pp. 65-80). Springer US.

Antin, J. F., Dingus, T. A., Hulse, M. C., & Wierwille, W. W. (1990). An evaluation of the effectiveness and efficiency of an automobile moving-map navigational display. *International Journal of Man-Machine Studies, 33*(5), 581-594.

Ardoin, W.V., & Ferris, T.K. (2016). Investigating redundant encoding methods for tactile messaging in multitask scenarios. *IEEE Transactions on Human-Machine Systems.*

Bacabac, R. G., Smit, T. H., Van Loon, J. J., Doulabi, B. Z., Helder, M., & Klein-Nulend, J. (2006). Bone cell response to high-frequency vibration stress: does the nucleus oscillate within the cytoplasm?. *The FASEB journal, 20*(7), 858-864.

Bennett, K.B., & Flach, J.M. (1992). Graphical displays: Implications for divided attention, focused attention, and problem solving. *Human Factors, 34*(5), 513-533.

BioHarness 3 Zephyr Technology Corporation. (2015). Retrieved from http://www. zephyranywhere.com/products/bioharness-3

Bodnar, A., Corbett, R. and Nekrasovski, D. (2004). Aroma: ambient awareness through olfaction in a messaging application. *Proceedings of ACM Inter-national Conference on Multimodal Interfaces (ICMI'04)*, 183-190.

Boucsein, W. (2012). *Electrodermal activity*. Springer Science & Business Media.

Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control.*Biological psychology*, *42*(3), 361-377.

Byers, J. C., Bittner, A. C., & Hill, S. G. (1989). Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary. *Advances in industrial ergonomics and safety I*, 481-485.

Carswell, C. M., Clarke, D., & Seales, W. B. (2005). Assessing mental workload during laparoscopic surgery. *Surgical innovation*, *12*(1), 80-90.

Carvalho, P. V., dos Santos, I. L., Gomes, J. O., Borges, M. R., & Guerlain, S. (2008). Human factors approach for evaluation and redesign of human–system interfaces of a nuclear power plant simulator. *Displays*, *29*(3), 273-284.

Casner, S., & Larkin, J. H. (1989). *Cognitive efficiency considerations for good graphic design* (No. AIP-81). Carnegie-Mellon Univ Pittsburgh Pa Artificial Intelligence And Psychology Project.

Chaudhury, H., Mahmood, A., & Valente, M. (2009). The effect of environmental design on reducing nursing errors and increasing efficiency in acute care settings: A review and analysis of the literature. *Environment and Behavior*.

Chen, H. Y., Santos, J., Graves, M., Kim, K., & Tan, H. Z. (2008). Tactor localization at the wrist. *Haptics: Perception, Devices and Scenarios*, 209-218. Springer Berlin Heidelberg.

Cholewiak, R. W., Brill, J. C., & Schwab, A. (2004). Vibrotactile localization on the abdomen: Effects of place and space. *Perception & Psychophysics, 66*(6), 970-987.

Cholewiak, R. W., & Collins, A. A. (1991). Sensory and physiological bases of touch. *The psychology of touch*, 23-60.

Cholewiak, R. W., & Collins, A. A. (2003). Vibrotactile localization on the arm: Effects of place, space, and age. *Perception & psychophysics*, *65*(7), 1058-1077.

Cinaz, B., Arnrich, B., La Marca, R., & Tröster, G. (2013). Monitoring of mental workload levels during an everyday life office-work scenario. *Personal and ubiquitous computing*, *17*(2), 229-239.

Clark, R. C., Nguyen, F., & Sweller, J. (2011). *Efficiency in learning: Evidence-based guidelines to manage cognitive load*. John Wiley & Sons.

Clark, R. E., Feldon, D., van Merriënboer, J. J., Yates, K., & Early, S. (2008). Cognitive task analysis. *Handbook of research on educational communications and technology*, *3*, 577-593.

Cummings, M. L., Myers, K., & Scott, S. D. (2006). Modified Cooper Harper evaluation tool for unmanned vehicle displays. *Proceedings of UVS Canada: Conference on Unmanned Vehicle Systems Canada*.

Darkow, D.J. and Marshak, W.P. (1998). In research of an objective metric for complex displays. *Proceedings of the Human Factors and Ergonomics Society42nd Annual Meeting*, 1361-1365.

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

Dinadis, N., & Vicente, K. J. (1996). Ecological interface design for a power plant feedwater subsystem. *Nuclear Science, IEEE Transactions on*, *43*(1), 266-277.

Durantin, G., Gagnon, J. F., Tremblay, S., & Dehais, F. (2014). Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behavioural brain research*, *259*, 16-23.

Eppler, M. J., & Mengis, J. (2004). The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. *The information society*, *20*(5), 325-344.

Estes, S. (2015). The Workload Curve Subjective Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*(7), 1174-1187.

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition & Emotion*, *6*(6), 409-434.

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: attentional control theory. *Emotion*, *7*(2), 336.

Fales, C. L., Barch, D. M., Burgess, G. C., Schaefer, A., Mennin, D. S., Gray, J. R., & Braver, T. S. (2008). Anxiety and cognitive efficiency: differential modulation of transient and sustained neural activity during a working memory task. *Cognitive, Affective, & Behavioral Neuroscience*, *8*(3), 239-253.

Ferris, T. K., & Sarter, N. (2011). Continuously informing vibrotactile displays in support of attention management and multitasking in anesthesiology. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *53*(6), 600-611.

Ferris, T.K., & Sarter, N. (2010). When content matters: The role of processing code in tactile display design. *IEEE Transactions on Haptics, 3*(3), 1-12.

Fiore, S. M., Scielzo, S., Jentsch, F., & Howard, M. L. (2006, October). Understanding performance and cognitive efficiency when training for X-ray security screening. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 50, No. 25, pp. 2610-2614). Sage Publications.

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics*, *24*(6), 631-645.

Green, B. G. (1977). The effect of skin temperature on vibrotactile sensitivity. *Perception & Psychophysics*, *21*(3), 243-248.

Grier, R., Wickens, C., Kaber, D., Strayer, D., Boehm-Davis, D., Trafton, J. G., & St. John, M. (2008). The Red-Line of Workload: Theory, Research, and Design. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 52*(18), 1204-1208.

Hart, S.G. (2006). NASA-task load index (NASA-TLX); 20 years later. *Proceedings of the human factors and ergonomics society annual meeting, 50*(9), 904-908.

Hart, S.G. and Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In Hancock, P. and Meshkati, N. (eds.), *Human mental workload*, North Holland B.V., Amsterdam, 139-183.

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

Heart Rate Variability (HRV) | Polar USA. (2016). Retrieved from http://support.polar.com/us-en/support/Heart_Rate_Variability__HRV_

Helton, W. S. (2004). Validation of a Short Stress State Questionnaire. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *48*(11), 1238-1242.

Hoggan, E. & Brewster, S. (2007). New parameters for tacton design. *CHI'07 extended abstracts on Human factors in computing systems*, 2417-2422.

Hsia, H. J. (1971). The information processing capacity of modality and channel performance. *AV Communication Review*, *19*(1), 51-75.

Hunt, R. E., & Newman, R. G. (1997). Medical knowledge overload: a disturbing trend for physicians. *Health care management review*, *22*(1), 70-75.

Ishii, H., Wisneski, C., Brave, S., Dahley, A., D., Gorbet, M., Ullmer, B., and Yarin, P. (1998). AmbientRoom: integrating ambient media with architectural space. *Conference Summary of CHI '98*.

Jafarinaimi, N., Forlizzi, J., Hurst, A., and Zimmerman, J. (2005). Breakaway: an ambient display to change human behavior. *CHI 2005*, 1945-1948.

Jones, L. A., & Sarter, N. B. (2008). Tactile displays: Guidance for their design and application. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(1), 90-111.

Jou, Y. T., Yenn, T. C., Lin, C. J., Yang, C. W., & Chiang, C. C. (2009). Evaluation of operators' mental workload of human–system interface automation in the advanced nuclear power plants. *Nuclear Engineering and Design*, *239*(11), 2537-2542.

Juris, M., & Velden, M. (1977). The pupillary response to mental overload. *Physiological Psychology*, *5*(4), 421-424.

Kaber, D. B., & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science*, *5*(2), 113-153.

Kalyuga, S., & Sweller, J. (2005). Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. *Educational Technology Research and Development*, *53*(3), 83-93.

Kappeler-Setz, C., Gravenhorst, F., Schumm, J., Arnrich, B., & Tröster, G. (2013).

Towards long term monitoring of electrodermal activity in daily life.*Personal and ubiquitous computing*, *17*(2), 261-271.

Karim, N., Hasan, J. A., & Syed, S. A. (2011). Heart rate variability: A review. *Basic and Science, 7*(1), 71-77.

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and instruction*, *12*(1), 1-10.

Kirsh, D. (2000). A few thoughts on cognitive overload. *Intellectica, 1*(30), 19-51.

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

Komiyama, O., Kawara, M., & De Laat, A. (2007). Ethnic differences regarding tactile and pain thresholds in the trigeminal region. *The Journal of Pain, 8*(4), 363-369.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive psychology*, *6*(2), 293-323.

Lau, N., Jamieson, G. A., Skraaning Jr, G., & Burns, C. M. (2008). Ecological Interface Design in the nuclear domain: An empirical evaluation of ecological displays for the secondary subsystems of a boiling water reactor plant simulator. *Nuclear Science, IEEE Transactions on*, *55*(6), 3597-3610.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(1), 50-80.

Lim, S. C., Kim, S. C., Hwang, J. H., & Kwon, D. S. (2010). Physical contact of devices: utilization of beats for interpersonal communication. *Haptics: generating and perceiving tangible sensations*, 348-353.

Lim, S. C., Kyung, K. U., & Kwon, D. S. (2012). Effect of frequency difference on sensitivity of beats perception. *Experimental brain research*, *216*(1), 11-19.

Liu, Y. C. (2001). Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveller information systems. *Ergonomics*, *44*(4), 425-442.

Liu, Y., Feyen, R., & Tsimhoni, O. (2006). Queueing Network-Model Human Processor (QN-MHP): A computational architecture for multitask performance in human-machine systems. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *13*(1), 37-70.

lo Storto, C. (2013). Evaluating ecommerce websites cognitive efficiency: an integrative framework based on data envelopment analysis. *Appl Ergon*, *44*(6), 1004-1014.

Luximon, A., & Goonetilleke, R. S. (2001). Simplified subjective workload assessment technique. *Ergonomics*, *44*(3), 229-243.

MacLean, K. E. (2009). Putting haptics into the ambience. *Haptics, IEEE Transactions on*, *2*(3), 123-135.

Makino, Y., & Maeno, T. (2013). Paired vibratory stimulation for haptic feedback. *Proceedings of the 4th Augmented Human International Conference*, 47-50. ACM.

Matthews, G., & Campbell, S. E. (2010). Dynamic relationships between stress states and working memory. *Cognition & Emotion*, *24*(2), 357-373.

Matthews, G., Reinerman-Jones, L. E., Barber, D. J., & Abich, J. (2015). The psychometrics of mental workload multiple measures are sensitive but divergent. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*(1), 125-143.

Mayer, R. E., & Massa, L. J. (2003). Three facets of visual and verbal learners: cognitive ability, cognitive style, and learning preference. *Journal of educational psychology*, *95*(4), 833.

McCracken, J. H., & Aldrich, T. B. (1984). *Analyses of selected LHX mission functions: Implications for operator workload and system automation goals* (No. ASI479-024-84). ANACAPA SCIENCES INC FORT RUCKER AL.

McFadden, D. (1998). Sex differences in the auditory system. *Developmental Neuropsychology*, *14*(2-3), 261-298.

Mehler, B., Reimer, B., & Coughlin, J. F. (2012). Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task an on-road study across three age groups. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 54*(3), 396-412.

Mehler, B., Reimer, B., & Wang, Y. (2011). A comparison of heart rate and heart rate variability indices in distinguishing single-task driving and driving under secondary cognitive workload. *Proceedings of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design,* 590-597.

Mehler, B., Reimer, B., Coughlin, J. F., & Dusek, J. A. (2009). Impact of Incremental Increases in Cognitive Workload on Physiological Arousal and Performance in Young Adult Drivers. *Transportation Research Record: Journal of the Transportation Research Board, 2138*(-1), 6-12.

Mehler, B., Reimer, B., Coughlin, J., & Dusek, J. (2009). Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record: Journal of the Transportation Research Board*, (2138), 6-12.

Miller, G. A. (1954). What is information measurement? *The American Psychologist, 8*, 3-11.

Miyake, S. (2001). Multivariate workload evaluation combining physiological and subjective measures. *International journal of psychophysiology*, *40*(3), 233-238.

Moacdieh, N., & Sarter, N. (2015). Display clutter a review of definitions and measurement techniques. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*(1), 61-100.

Mumaw, R. J., Roth, E. M., Vicente, K. J., & Burns, C. M. (2000). There is more to monitoring a nuclear power plant than meets the eye. *Human factors: The journal of the human factors and ergonomics society*, *42*(1), 36-55.

Murray, N. P., & Janelle, C. M. (2003). Anxiety and performance: A visual search examination of the processing efficiency theory. *Journal of Sport and Exercise Psychology*, *25*(2), 171-187.

Nachreiner, F. (1995). Standards for ergonomics principles relating to the design of work systems and to mental workload. *Applied Ergonomics*, *26*(4), 259-263.

Nemenman, I., Lewen, G. D., Bialek, W., & van Steveninck, R. R. D. R. (2008). Neural coding of natural stimuli: information at sub-millisecond resolution. *PLoS Comput Biol*, *4*(3), e1000025.

O'Donnell, C. R., and Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman, & J. P. Thomas *(eds.), Handbook of perception and human performance*: Vol. II. Cognitive processes and performance (pp. 42.1-42.29). New York: Wiley.

Ogden, G. D., Levine, J. M., & Eisner, E. J. (1979). Measurement of workload by secondary tasks. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *21*(5), 529-548.

Oster, G. (1973). Auditory beats in the brain. *Scientific American*, *229*(4), 94-102.

Oviatt, S. (1996). Multimodal interfaces for dynamic interactive maps. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 95-102.

Paas, F. G., & Van Merriënboer, J. J. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *35*(4), 737-743.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational psychologist*, *38*(1), 63-71.

Parasuraman, R., and Caggiano, D. (2002). Mental Workload. in Ramachandran, V.S. (ed.), *Encyclopedia of the Human Brain*, Vol. 3, Academic Press, San Diego, CA, 17–27.

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

Penfield, W., & Welch, K. (1951). The supplementary motor area of the cerebral cortex: a clinical and experimental study. *Archives of Neurology and Psychiatry*, *66*(3), 289.

Piateski, E., & Jones, L. (2005). Vibrotactile pattern recognition on the arm and torso. *Proceedings of the First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 90-95. IEEE.

Pollack, I. (1952). The information of elementary auditory displays. *The Journal of the Acoustical Society of America*, *24*(6), 745-749.

Pollack, I. (1953). The information of elementary auditory displays. II. *The Journal of the Acoustical Society of America*, *25*(4), 765-769.

Polson, P. G., Lewis, C., Rieman, J., & Wharton, C. (1992). Cognitive walkthroughs: a method for theory-based evaluation of user interfaces. *International Journal of man-machine studies*, *36*(5), 741-773.

Pongrac, H. (2008). Vibrotactile perception: examining the coding of vibrations and the just noticeable difference under various conditions. *Multimedia systems*,*13*(4), 297-307.

Pousman, Z. and Stasko, J. (2006). A taxonomy of ambient information systems: Four patterns of design. *Proceedings of Advanced Visual Interfaces*, 67–74.

Rasmussen, J. (1999). Ecological interface design for reliable human-machine systems. *The International Journal of Aviation Psychology*, *9*(3), 203-223.

Reid, G. B., & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. *Advances in psychology*, *52*, 185-218.

Reid, G. B., Eggemeier, F. T., & Shingledecker, C. A. (1982). *Subjective workload assessment technique*. Air Force Flight Test Center Edwards AFB CA.

Rodriguez Paras, C. (2015). *Exploring physiological measures for prediction and identification of the redline of cognitive workload*. Master's thesis, Texas A&M University.

Rodriguez Paras, C., Yang, S., Tippey, K., & Ferris, T. K. (2015). Physiological Indicators of the Cognitive Redline. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*,59(1), 637-641.

Ross, T., & Burnett, G. (2001). Evaluating the human–machine interface to vehicle navigation systems as an example of ubiquitous computing. *International Journal of Human-Computer Studies*, *55*(4), 661-674.

Rubio, S., Díaz, E., Martín, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology*, *53*(1), 61-86.

Ryu, K., & Myung, R. (2005). Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. *International Journal of Industrial Ergonomics*, *35*(11), 991-1009.

Salden, R. J., Paas, F., Broers, N. J., & Van Merriënboer, J. J. (2004). Mental effort and performance as determinants for the dynamic selection of learning tasks in air traffic control training. *Instructional science*, *32*(1-2), 153-172.

Sanderson, P., Anderson, J., & Watson, M. (2000, December). Extending ecological interface design to auditory displays. In *Proceedings of the 10th Australasian Conference on Computer-Human Interaction* (pp. 259-266).

Sarter, N. B. (2002). Multimodal information presentation in support of human-automation communication and coordination. *Advances in human performance and cognitive engineering research*, *2*, 13-36.

Sarter, N. B. (2006). Multimodal information presentation: Design guidance and research challenges. *International journal of industrial ergonomics*, *36*(5), 439-445.

Schraagen, J. M., Chipman, S. F., & Shalin, V. L. (Eds.). (2000). *Cognitive task analysis*. Psychology Press.

Seagull, F.J., Wickens, C., & Loeb, R.G. (2001). When is less more? Attention and workload in auditory, visual, and redundant patient-monitoring conditions.

Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting, Santa Monica, CA, 1395-1399.

Shannon, C. E., & Weaver, W. (1949). The mathematical theory of communications. Urbana: University of Illinois Press.

Shannon,C.E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 379-423, 623-656.

Sharp, T. D., & Helmicki, A. J. (1998). The application of the ecological interface design approach to neonatal intensive care medicine. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *42*(3), 350-354.

Sherrick, C. E., & Cholewiak, R. W. (1986). Cutaneous sensitivity. *Handbook of perception and human performance*, *1*, 1-12. *Sensory Processes and Perception*. John Wiley & Sons.

Shi, Y., Ruiz, N., Taib, R., Choi, E., & Chen, F. (2007, April). Galvanic skin response (GSR) as an index of cognitive load. In *CHI'07 extended abstracts on Human factors in computing systems* (pp. 2651-2656). ACM.

Shimmer3 GSR Unit. (2016). Retrieved from http://www.shimmersensing.com/shop /shimmer3-wireless-gsr-sensor

Simon, H.A. (1999). *The Sciences of the Artificial*, 3rd ed. Cambridge, MA: MIT Press.

Taelman, J., Vandeput, S., Spaepen, A., & Van Huffel, S. (2009). Influence of mental stress on heart rate and heart rate variability. In *4th European conference of the international federation for medical and biological engineering* (pp. 1366-1369). Springer Berlin Heidelberg.

Tan, H. Z., Reed, C. M., & Durlach, N. I. (2010). Optimum information transfer rates for communication through haptic and other sensory modalities. *Haptics, IEEE Transactions on*, *3*(2), 98-108.

Tan, Z. C., Reimer, B., Mehler, B., & Coughlin, J. F. (2011). Detection of elevated states of cognitive demand in drivers in a naturalistic driving environment. In *Proceedings* *of the 2nd Annual International Conference on Advanced Topics in Artificial Intelligence (ATAI 2011)* (pp. 68-73).

Task Force of the European Society of Cardiology. (1996). Heart rate variability standards of measurement, physiological interpretation, and clinical use. *Eur Heart J*, *17*, 354-381.

Tobias, J. V. (1965). Consistency of sex differences in binaural-beat perception. *International Journal of Audiology*, *4*(2), 179-182

Tsang, P. S., & Velazquez, V. L. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics*, *39*(3), 358-381.

Vallbo, A. B., & Johansson, R. S. (1984). Properties of cutaneous mechanoreceptors in the human hand related to touch sensation. *Hum Neurobiol*, *3*(1), 3-14

Vicente, K. J. (1992). Memory recall in a process control system: A measure of expertise and display effectiveness. *Memory & Cognition*, *20*(4), 356-373.

Vicente, K. J. (2002). Ecological interface design: Progress and challenges. *Human Factors: The Journal of the Human Factors and Ergonomics Society*,*44*(1), 62-78.

Vicente, K. J., & Rasmussen, J. (1992). Ecological interface design: Theoretical foundations. *Systems, Man and Cybernetics, IEEE Transactions on*, *22*(4), 589-606.

Vidulich, M. A., & Tsang, P. S. (2012). Mental workload and situation awareness. *Handbook of Human Factors and Ergonomics*, 243-273.

Villarejo, M. V., Zapirain, B. G., & Zorrilla, A. M. (2012). A stress sensor based on Galvanic Skin Response (GSR) controlled by ZigBee. *Sensors*,*12*(5), 6075-6101.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*(6984), 748-751.

Vollmann, T. E. (1991). Cutting the Gordian knot of misguided performance measurement. *Industrial Management & Data Systems*, *91*(1), 24-26.

Wahbeh, H., Calabrese, C., & Zwickey, H. (2007). Binaural beat technology in humans: a pilot study to assess psychologic and physiologic effects. *The Journal of Alternative and Complementary Medicine*, *13*(1), 25-32.

Walker, B. N., Stanley, R. M., Iyer, N., Simpson, B. D., & Brungart, D. S. (2005). Evaluation of bone-conduction headsets for use in multitalker communication environments. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 49*(17), 1615-1619. SAGE Publications.

Watson, M. O., & Sanderson, P. M. (2007). Designing for attention with sound: challenges and extensions to ecological interface design. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *49*(2), 331-346.

Watson, M.O. & Sanderson, P.M. (2004). Sonification helps eyes-free respiratory monitoring and task timesharing. *Human Factors,46*(3), 497-517.

Weiser, M., & Brown, J. S. (1996). Designing calm technology. *PowerGrid Journal, 1*(1), 75-85.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Sciences,* 3, 159-177.

Wickens, C. D., Hollands, J. G., Parasuraman, R., & Banbury, S. (2012). *Engineering Psychology and Human Performance (4th Edition), Pearson*.

Wickens, C. D., Hollands, J.G., Banbury, S., and Parasuraman, R. *Engineering psychology & human performance*. Psychology Press, 2015.

Wickens, C., Prinet, J., Hutchins, S., Sarter, N., & Sebok, A. (2011). Auditory-Visual Redundancy in Vehicle Control Interruptions Two Meta-analyses. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 55*(1), 1155-1159.

Wickens, C.D., Hollands, J.G., Parasuraman, R., & Banbury, S. (2012). *Engineering Psychology and Human Performance (4th Edition)*. New York, NY: Pearson.

Wickens, C.D., Lee, J.D., Liu, Y., & Gordon-Becker, S. (2004). *Introduction to Human Factors Engineering (2nd Edition)*. Pearson Education, Inc., NJ.

Wiebe, E. N., Roberts, E., & Behrend, T. S. (2010). An examination of two mental workload measurement approaches to understanding multimedia learning. *Computers in Human Behavior, 26*(3), 474-481.

Wilson, G. F. (2002). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation Psychology*, *12*(1), 3-18.

Wilson, M., Smith, N. C., Chattington, M., Ford, M., & Marple-Horvat, D. E. (2006). The role of effort in moderating the anxiety–performance relationship: Testing the prediction of processing efficiency theory in simulated rally driving. *Journal of Sports Sciences*, *24*(11), 1223-1233.

Wolfe, J. M., Dluender, K. R., & Levi, D.M. (2011). *Sensation & Perception, Third Edition*. MA: Sinauer Associates, Inc.

Woods, D. D. (1995). The alarm problem and directed attention in dynamic fault management. *Ergonomics*, *38*(11), 2371-2393.

Woods, D. D., Patterson, E. S., & Roth, E. M. (2002). Can we ever escape from data overload? A cognitive systems diagnosis. *Cognition, Technology & Work*, *4*(1), 22-36.

Woods, D. D., Wise, J. A., & Hanes, L. F. (1981). An evaluation of nuclear power plant safety parameter display systems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 25*(1), 110-114.

Xie, B., & Salvendy, G. (2000). Review and reappraisal of modelling and predicting mental workload in single-and multi-task environments. *Work & stress, 14*(1), 74-99.

Yang, S., Nevins, L., & Ferris, T. K. (2015). Investigating Redundant Multimodal Speedometer Displays for Supporting Concurrent Lane and Speed Tracking. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, *59*(1), 1578-1582.

Yang, S., Shukla, K., & Ferris, T. K. (2012). "Cognitive Efficiency" in Display Media: A First Investigation of Basic Signal Dimensions. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 1371-1375.

Yang, S., Tippey, K., & Ferris, T. K. (2014). Exploring the Emergent Perception of Haptic Beats from Paired Vibrotactile Presentation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58*(1), 1716-1720.

Yang, S., You, N. Y., & Ferris, T. K. (2013). Supporting drivers in concurrent lane and speed tracking tasks with novel visual, auditory, and tactile speedometer displays. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 57(1), 1918-1922. SAGE Publications.

APPENDIX A

Subjective Self-reported Measure			
Measures	Dependent Variables	Advantages	Limitations
NASA Task Load Index (NASA TLX; Hart & Staveland, 1988)	NASA TLX has six scales: • mental demand • physical demand • temporal demand • performance • effort • frustration	• No interruption on primary task performance • Relatively easy to use • NASA TLX scale is reliably sensitive to experimental manipulation (Hart & Staveland, 1988) • Provides task information that is not available from SWAT (Rubio, Díaz, Martín, & Puente, 2004)	• Less sensitive to task combinations • Less sensitive to output modality (i.e., speed and manual) manipulations (Hart & Staveland, 1988)
Raw Task Load Index (RTLX; Byers, 1989)	The average of all six NASA TLX scales (Byers, 1898)		
Subjective Workload Assessment Technique(SWAT; Reid & Nygren, 1988)	SWAT has three scales: • time load • mental effort load • psychological stress load	• Ease to use • Not intrusive • Low cost • High validity • High sensitivity to workload variations	• Low sensitivity to low mental workload • It requires a time-consuming card sorting pre-task procedure (Luximon & Goonetilleke, 2001)
Workload Profile Method (Tsang & Velazquez, 1996)	WP has eight scales: • perception • response selection and execution • spatial processing • verbal processing • visual processing • auditory processing • manual output • speech output	• It shows the highest sensitivity to sources of workload, comparing to NASA-TLX and SWAT (Rubio, Díaz, Martín, & Puente, 2004) • It has higher power of diagnosticity than NASA-TLX and SWAT • WP is sensitive to task combination • WP is appropriate when we need to analyze cognitive demands or attention resources demanded by a particular task (Rubio, Díaz, Martín, & Puente, 2004)	

Continued

Measures	Dependent Variables	Advantages	Limitations
Simplified Workload Assessment Technique (Luximon & Goonetilleke, 2001)	Five variables: • A_{SWAT} • D_{SWAT} • W_0 • W_1 • PC_c	• A_{SWAT} has the highest sensitivity • The sensitivity of D_{SWAT} is still better than SWAT, although not good as A_{SWAT} • Less time cost compared to SWAT	• Further validation may be needed with tasks of varying workload to confirm the findings
Modified Cooper-Harper Scale (MCH; Wierwille & Casali, 1983)		• The ability to analyze a remote operator's ability to effectively and efficiently complete and manage higher-level system tasks (Cummings, Myers, & Scott, 2006) • Evaluate how well operators can achieve the goal of higher level tasking • Illustrate relationship between display and cognition process (Cummings, Myers, & Scott, 2006)	• Variation in measures can be significant • Subjective measures may not necessarily guide interface design • The numerical result should be interpreted in light of user's comment (Cummings, Myers, & Scott, 2006)
Rating Scale Mental Effort (RSME; Zijlstra & Meijman, 1989; Zijilstra & Van Doon, 1985)			
Bedford Scale (Roscoe, 1987; Roscoe & Ellis, 1990)			

Performance Measure			
Measures	Dependent Variables	Advantages	Limitations
Primary-task Performance (O' Donnell & Eggemeier, 1986)	• The performance on the system of primary interests, such as the number of error and speed of performance	• It's a direct measure of mental workload	• The 'underload' condition can reduce the sensitivity of the measures • Two primary tasks to be compared may differ in how they are measured • Sometimes it's impossible to obtain good measures of primary-task performance • The difference of two primary task performance may be caused by the difference of the limitation of data
Secondary-task Performance (Ogden, Levine, & Eisner, 1979)	The measured variables of secondary-task performance usually include: • reaction time • mental arithmetic • self-adaptive tracking • monitoring	• Noninterference • Ease to use • Self-pacing • Continuity of scoring • Compatibility with the primary task • High sensitivity • Good representativeness	• Secondary task index is not always sensitive • Secondary task may interfere with and disrupt the primary-task performance (Wickens, Hollands, Parasuraman, & Banbury, 2012) • Selection of secondary task could be a complex process

<table>
<tr><th colspan="4">Physiological Measure</th></tr>
<tr><td>Measures
(De Waard, 1996)</td><td>Dependent Variables</td><td>Advantages</td><td>Limitations</td></tr>
<tr><td>Cardiac Functions
Electrocardiogram(ECG)</td><td>• Heart Rate
• Heart Rate Variability
• T-wave</td><td rowspan="10">The general advantages of physiological measures include (De Waard, 1996):

• they do not require an overt response by the operator;
• most physiological variables can be recorded continuously;
• they are unobtrusive due to miniaturization.</td><td rowspan="10">The general limitations of physiological measures are summarized at several points below

• Individual differences, such as age, could affect physiological measures;
• Environmental factors, such an ambient light and temperature, can affect some physiological measures. For example, strong ambient light can affect the pupil diameter;
• Data of physiological measure could still be noisy;
• Wearing sophisticated physiological devices might discomfort the participants
•</td></tr>
<tr><td>Electroencephalogram (EEG)</td><td>• Alpha waves
• Beta waves
• Delta waves
• Theta waves</td></tr>
<tr><td>Eye fixations
Electroculogram (EOG)</td><td>• Fixation time
• Range of saccadic</td></tr>
<tr><td>Pupil diameter</td><td>• Size of Pupil diameter</td></tr>
<tr><td>Endogenous Eye Blinks</td><td>• Eye blink rate
• Blink duration
• Eye blink latency</td></tr>
<tr><td>Blood Pressure</td><td>• Blood pressure variability</td></tr>
<tr><td>Respiration</td><td>• Respiration rate</td></tr>
<tr><td>Electrodermal Activity (EDA)</td><td>• Skin conductance level
• Skin conductance response</td></tr>
<tr><td>Event-related Potentials (ERP)</td><td></td></tr>
<tr><td>Hormone Levels</td><td></td></tr>
</table>