Computational Modeling and Experimental Research on Touchscreen Gestures, Audio/Speech Interaction, and Driving

by

Heejin Jeong

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Operations Engineering)
in the University of Michigan
2018

Doctoral Committee
Professor Yili Liu, Chair
Professor Melissa Gross
Associate Professor Bernard J. Martin
Professor Nadine B. Sarter

Heejin Jeong

heejinj@umich.edu

ORCID iD: 0000-0002-0122-532X

To my family

Acknowledgements

First and foremost, I would like to express my sincere and the greatest thanks to my advisor, Dr. Yili Liu, for his invaluable advice and guidance on the cognitive modeling research throughout the dissertation stages. He has been always listening and understanding my words with patience, while always encouraging me to be creative about my research. His support and advice have also helped me develop a broader picture of my life as a researcher and what I can expect after graduation.

I would also like to appreciate all members of my dissertation committee, Dr. Bernard Martin, Dr. Melissa Gross, and Dr. Nadine Sarter, for their time and priceless advice during the dissertation process. I would also like to thank my master degree advisor, Dr. Paul Green for his great mentorship on transportation human factors research; Charles Woolley for being my great teaching advisor through the last six semesters. I also want to thank Dr. Fred Feng for his help, particularly in getting to know the simulation model and software; Dr. Yifan Chen for his professional guidance and support; Dr. Shi Cao, Sang Won Lee, and Aravindh Sabarish Ramakrishnan for assisting me with the research. I extend my gratitude to the many professors, my student colleagues, and staff members in the IOE department for their help and support.

This dissertation work and my graduate studies were supported by funds from multiple sources: a research project collaborating with Ford Motor Company, the fellowship and graduate student instructor positions provided by the Department of Industrial and Operations Engineering at the University of Michigan, a Rackham Graduate Student Research Grant and four-time of Rackham Conference Travel Grants from the Horace H. Rackham School of Graduate Studies at the University of Michigan, a National Science Foundation research project under the guidance of Dr. Paul Green, and the M. Scheller Fellowship by the College of Engineering at the University of Michigan.

Finally, I wish to thank my wonderful family in the U.S. and my home country. I am especially grateful to my wife, Miyoung, for her unconditional support with love, and to our son, William, for making our lives more meaningful.

Table of Contents

Lists of Tables

Lists of Figures

Abstract

As humans are exposed to rapidly evolving complex systems, there are growing needs for humans and systems to use multiple communication modalities such as auditory, vocal (or speech), gesture, or visual channels; thus, it is important to evaluate multimodal human-machine interactions in multitasking conditions so as to improve human performance and safety. However, traditional methods of evaluating human performance and safety rely on experimental settings using human subjects which require costly and time-consuming efforts to conduct. To minimize the limitations from the use of traditional usability tests, digital human models are often developed and used, and they also help us better understand underlying human mental processes to effectively improve safety and avoid mental overload. In this regard, I have combined computational cognitive modeling and experimental methods to study mental processes and identify differences in human performance/workload in various conditions, through this dissertation research.

The computational cognitive models were implemented by extending the Queuing Network-Model Human Processor (QN-MHP) Architecture that enables simulation of human multi-task behaviors and multimodal interactions in human-machine systems. Three experiments were conducted to investigate human behaviors in multimodal and multitasking scenarios, combining the following three specific research aims that are to understand: (1) how humans use their finger movements to input information on touchscreen devices (i.e., touchscreen gestures), (2) how humans use auditory/vocal signals to interact with the machines (i.e., audio/speech interaction), and (3) how humans drive vehicles (i.e., driving controls). Future research applications of computational modeling and experimental research are also discussed.

Scientifically, the results of this dissertation research make significant contributions to our better understanding of the nature of touchscreen gestures, audio/speech interaction, and driving controls in human-machine systems and whether they benefit or jeopardize human performance and safety in the multimodal and concurrent task environments. Moreover, in contrast to the previous models for multitasking scenarios mainly focusing on the visual processes, this study develops quantitative models of the combined effects of auditory, tactile, and visual factors on multitasking performance. From the practical impact perspective, the modeling work conducted in this research may help multimodal interface designers minimize the limitations of traditional usability tests and make quick design comparisons, less constrained by other time-consuming factors, such as developing prototypes and running human subjects. Furthermore, the research conducted in this dissertation may help identify which elements in the multimodal and multitasking scenarios increase workload and completion time, which can be used to reduce the number of accidents and injuries caused by distraction.

Chapter 1.

Introduction

1.1. Overview

As the amount of information that humans and machines exchange with each other increases, there are increasing demands for humans and machines to use multiple communication modalities such as auditory, vocal (or speech), gesture, or visual channels. Multimodal interfaces are among the most frequently used human-machine interfaces in many complex and data-rich environments, especially in multi-task situations such as automotive, aerospace, and medical fields. Multimodal interfaces have advantages over unimodal interfaces, including providing more bandwidth to the communication and supporting functions such as timesharing, redundancy, synergy, disambiguation (Sarter, 2006). However, the use of multimodalities has some constraints: the existence of cross-modal links in attention in the form of (1) modality attention, (2) the modality shifting effect, and (3) cross-modal spatial and temporal links (for more details, see Spence & Driver, 1997). Other literatures also emphasized a drawback of multimodal interfaces, which is the mental overload derived from the information with too many channels in an interface (Jaimes & Sebe, 2007; Jokinen & Raike, 2003). These may increase users' cognitive workload and reduce their performance, which eventually could lead to higher chances of operating errors and accidents (Cook, Potter, Woods, & McDonald, 1991). In order to minimize the limitations, it is necessary to evaluate the efficiency of multimodal interfaces in multitasking environments.

In this dissertation research, I conducted experimental studies and developed corresponding computational models to investigate human behaviors in multimodal and multitasking scenarios. Specific goals of my dissertation include investigating:

1) How humans use their finger movements to input information on touchscreen devices (i.e., touchscreen gestures)
2) How humans use auditory/vocal signals to interact with machines (i.e., audio/speech interaction)
3) How humans drive vehicles (i.e., driving control)

Touchscreen gestures: Touchscreen devices have become widely used and the corresponding use of finger input gestures has become an important part of our everyday lives (Minsky, 1984; Poupyrev & Maruyama, 2003). Due to numerous advantages of touchscreens from the human factors perspective, including (1) intuitive and direct manipulation and (2) contextual presentation of controls, a large number of touchscreen user interfaces have been developed and become ubiquitous in many application domains including driving, healthcare, education, and public access information systems (Shneideman, 1991; Viviani & Calil, 2015). In this respect, it is important to evaluate the touchscreen interfaces with finger gestures and predict human performance in using touchscreens.

Although the terms vary slightly, touchscreen finger gestures consist of two categories: single-touch (e.g., clicking (or pressing), swiping (or flicking), scrolling (or dragging)) and multi-touch (e.g., pinching (or zooming-out), and spreading (or zooming-in)). Several touchscreen gesture studies have been conducted in the domains of human-computer interactions and human

factors (e.g., Kobayashi et al, 2011; Jennings, Ryser, & Drews, 2013; Findlater, Froehlich, Fattal, Wobbrock, & Dastyar, 2013; Hoggan et al., 2013) and they were mostly experimental studies using human subjects.

Audio/Speech interaction: Many spoken dialogue systems have been developed as both research prototypes and commercial applications (Glass, 1999). One of the successful uses of spoken dialogue systems is automotive navigation systems with conversational speech interactions. Recent in-vehicle navigation systems are equipped with not only guidance generation with synthesized voice, but also with operation acceptance with speech recognition (Kono, Yano, & Sasajima, 1999). Much research has examined the benefits of hands-free speech (or voice) interactions in the automotive fields (Barón & Green, 2006; Gärtner, König, & Wittig, 2001). The use of audio/speech interaction results in less mental taxation, fewer glances off the road compared to manual data entry, and better driving performance especially in more complex tasks such as navigation and phone dialing. While many human dialogue phenomena have been successfully modeled (e.g., Mann, Moore, & Levin, 1977; Traum, Swartout, Gratch, & Marsella, 2008), there are few modeling studies using cognitive architectures and quantitative prediction methods for evaluating spoken dialogue systems.

Driving control: In driving, mainly two types of controls are required: (1) lateral (also called steering) and (2) longitudinal (also called speed) controls. For both controls, drivers must continuously adjust or maintain their steering and speed while looking at the road, monitoring the speedometer, operating the in-vehicle systems (e.g., navigation or radio), and making critical decisions regarding steering and speed. In order to support quantitative analysis of such complex behavior of driving controls in multi-tasking environments, it is essential to develop computational models to predict human performance.

1.2. Research Objectives

This dissertation research utilizes both computational modeling and experimental methods to investigate human behaviors in multimodal and multitasking scenarios (including touchscreen gestures, audio/speech interaction, and driving). The overarching objectives of the research are to:

(1) Develop computational models of human-machine multimodal interactions that can predict human performance and workload (Phase I).
(2) Conduct experimental studies to examine how humans interact with machines (Phase II).
(3) Validate the models with human performance data (Phase III).

In Phase I, a task-independent cognitive architecture, the Queuing Network-Model Human Processor (QN-MHP; Liu, Feyen, & Tsimhoni, 2006), is used to model and implement touchscreen gestures and audio/speech interaction in a variety of conditions including driving. In Phase II, laboratory experiments are conducted to obtain quantitative evidence to evaluate the models developed in the first phase. In Phase III, the experimental data from the second phase (and/or from the literature) are used to determine whether the models are an accurate representation of the real system.

1.3. Queueing Network – Model Human Processor

1.3.1. Overview

The Queuing Network – Model Human Processor (QN-MHP; Liu et al., 2006) is a computational cognitive architecture that integrates the mathematical framework of queueing network theory with the Model Human Processor (Card, Moran, & Newell, 1983) model. The

QN-MHP is composed of three subnetworks including perceptual, cognitive, and motor subnetworks (See Figure 1.1). Each subnetwork has multiple servers and they are connected by routes, representing information flow through human brain and body to conduct the given tasks. Each subnetwork is composed of servers that perform distinct functions. The selection of the servers and the connection between the servers are developed on the basis of related multiple findings in the psychology and neuroscience (Liu et al., 2006). Examples of the process that occurs in the perceptual subnetwork include to look at objects (i.e., visual perception) or to listen to sounds (i.e., auditory perception). In the cognitive subnetwork, the process includes to remember the objects or the sounds, and to compare perceived signals to expected signals. Examples in the motor subnetwork are to reach with hands (i.e., manual response) or to say words (i.e., speech/verbal response).

Visual inputs/entities enter the QN system through the common visual processing server (Server 1) and these incoming entities are transmitted in parallel routes; one is moved to the visual recognition server (Server 2) and another is moved to the visual location server (Server 3). These entities from the two visual routes are integrated at Server 4 (the visual recognition and location integration server). On the other hand, auditory entities enter through Server 5 and the entities travel in parallel routes; one moves to the auditory recognition server (Server 6) and another goes to the auditory server (Server 7). As the visual entities do, these auditory entities from the two auditory routes are integrated at Server 8 (the auditory recognition and location integration server).

The entities collected in Servers 4 and 8 travel to the cognitive subnetwork by entering through Server A (Visuospatial sketchpad) and Server B (Phonological loop). Out of the visual entities, the entities related to color/orientation information move to Server A, whereas text-

related visual entities move to Server B. All auditory entities including voice messages and sounds move to Server B. Server C is the central executive server, where the desired touchscreen gesture performance (e.g., movement time), speech (or voice) interaction performance (e.g., recognition processing time), and driving performance (e.g., steering wheel angle and pedal acceleration) are calculated. Server F is the complex cognitive function, where more complex mental processing is conducted, such as comparison and computation. Server C communicates with Server F, and then transmits entities to the motor subnetwork.

The motor subnetwork contains five servers. Once entities are transmitted to the motor subnetwork, motor programs and long-term procedural information are retrieved at Server W (the motor program retrieval server) and assembled at Server Y (the motor program assembling and error detecting server). After that, the primary motor server (Server Z) determines the level of exerted forces for the motor controls such as foot, hand, eyes, and mouth controls. Then, the neural signals are transmitted from Server Z to the foot, hand, eyes, and mouth servers. These servers execute (1) foot movements for pedal operation, (2) hand movements for steering operation and touchscreen gesture operation, (3) eye movements for looking operation, (4) and mouth controls for speaking operation, respectively. When the body segment (i.e., foot, hand, eyes, and mouth) starts to work, Server X (the sensorimotor integration server) collects motor information from Server Z and sensory information (e.g., visual, auditory, tactile feedback) from the body segment in real time, and then relays them to Server C as well as Server Y.

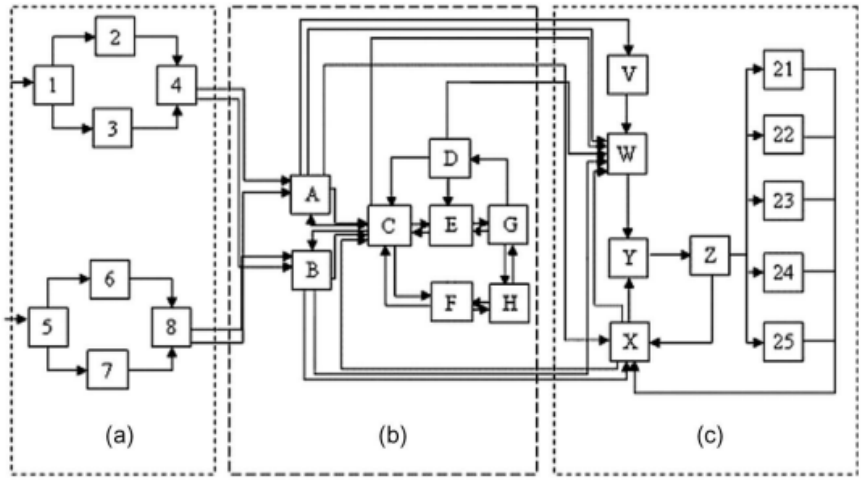
Here are three major reasons to support the QN architecture for its capability in modeling human behaviors in multitasking and multimodal scenarios:

First, the QN-MHP contains a hybrid system that enables both serial and parallel cognitive process modeling; in contrast, other cognitive architectures work only in either serial

(e.g., ACT-R) or parallel (e.g. EPIC). The QN-MHP allows more than one server to act in parallel or in serial simultaneously. Thus, it is possible to model human performance in multi-task scenarios represented as multiple flows of entities, such as driving performance under multi-task conditions.

Second, multimodal tasks can be implemented in the QN-MHP architecture, using the functions of servers in the perceptual and motor subnetworks. For example, the interaction of perception and motor controls can be modeled through the QN architecture: multiple sensory channels (e.g., visual or auditory; Severs 1-4 and 5-8) and responding methods (e.g., manual or speech; the hands and mouth servers).

Third, the QN-MHP can be used to model individual difference in human performance. In other words, the same task could be processed through different routes and servers, depending on the human capacity level. For example, Servers C (the central executive server) and/or F (the complex cognitive function server) can be selected and used for the particular cases by their assumptions. For those who are familiar with a certain task due to its repetition or learning effect, Server C is used to implement relatively simple mental processing. Server F can be used to implement more complex mental processing for the unexperienced persons or novices for the certain task.



Perceptual Subnetwork

1. Common visual processing (eyes, lateral geniculate nucleus, superior colliculus, primary and secondary visual cortex)
2. Visual recognition (dorsal system)
3. Visual location (ventral system)
4. Visual recognition and location integration (distributed parallel area including the connections among V3 , V4 and V5, superior frontal sulcus, and inferior frontal gyrus)
5. Common auditory processing (middle and inner ear)
6. Auditory recognition (area from dorsal and ventral cochlear nuclei to the inferior colliculus)
7. Auditory location (area from ventral cochlear nucleus to the superior olivary complex)
8. Auditory recognition and location integration (primary auditory cortex and planum temporale)

Cognitive Subnetwork

A. Visuospatial sketchpad (right-hemisphere posterior parietal cortex)
B. Phonological loop (left-hemisphere posterior parietal cortex)
C. Central executive (dorsolateral prefrontal cortex (DLPFC), anterior-dorsal prefrontal cortex (ADPFC) and middle frontal gyrus (GFm))
D. Long-term procedural memory (striatal and cerebellar systems)
E. Performance monitor (anterior cingulate cortex)
F. Complex cognitive function: decision, calculation, anticipation of stimulus in simple reaction etc. (intraparietal sulcus (IPS), the superior frontal gyrus (SFS), the inferior frontal gyrus (GFi), the inferior parietal cortex and the ventrolateral frontal cortex, the intraparietal sulcus and the superior parietal gyrus)
G. Goal initiation (orbitofrontal region and amygdala complex)
H. Long-term declarative & spatial memory (hippocampus and diencephalons)

Motor Subnetwork

V. Sensorimotor integration (premotor cortex)
W. Motor program retrieval (basal ganglia)
X. Feedback information collection (somosensoy cortex)
Y. Motor program assembling and error detecting (supplementary motor area (SMA) and the pre-SMA)
Z. Sending information to body parts (primary motor cortex)
21-25: Body parts: eye, mouth, left hand, right hand, foot

Figure 1.1 – General structure of the QN-MHP cognitive architecture (Liu, Feyen, & Tsimhoni, 2006; Wu & Liu, 2008)

Over the past years, the QN-MHP architecture has been successfully applied to a wide range of tasks such as simple and choice reaction time task (Feyen & Liu, 2001), map reading (Liu et al., 2006), transcription typing (Wu & Liu, 2008), visual search (Lim & Liu, 2004; Feng & Liu, 2013), and driving controls (Tsimhoni & Liu, 2003; Zhao & Wu, 2013; Bi, Gan, Shang, & Liu, 2012). Fuller and her colleagues (2012) integrated the QN-MHP model with physical digital human models to simulate both physical and cognitive performance. However, there have been no attempts for touchscreen gesture-, audio/speech-, and curve-driving-related modeling using the QN-MHP framework.

1.3.2. Model implementation in the QN-MHP

To implement human performance models in the QN-MHP architecture, it is required to (1) conduct a task analysis and (2) use operators based on the result of the task analysis (or develop new operators if they are needed but do not exist yet), and (3) use (or develop) digital device mockups, especially for human-machine interface (HMI) evaluations. The task analysis is conducted using the NGOMSL (Natural Goals, Operators, Methods, and Selection rules Language)-style task description (Kieras, 1999). The operator refers to the most elementary component of the task. Each operator is set with several parameters for specifying the task. To develop the operator, it is required to have logically solid assumptions and quantitative models (See the details in Liu et al., 2006; Chapters 3, 5, 7, 8 of this dissertation).

1.4. Contributions

1.4.1. Practical significance

In contrast to the high cost and lengthy test time associated with conventional usability evaluations using human subjects, computational models of human performance have significant benefits. The modeling work conducted in this research will help multimodal interface designers minimize the limitations of traditional usability tests and make quick design comparisons, less constrained by other time-consuming factors, such as developing prototypes and testing human subjects.

Furthermore, one of the less explored but important areas in the human factors engineering field is a development of computational models to predict driver behaviors in multi-task and multimodal scenarios. The research conducted in this dissertation may help understand

which elements in the multimodal and multitasking scenarios increase workload and completion time, which can be used to reduce the number of accidents and injuries caused by distraction.

1.4.2. Scientific significance

The research described in this dissertation makes important contributions to our better understanding of the nature of gesture and speech interactions in human-machine systems and whether they benefit or jeopardize human performance and safety in the multimodal and concurrent task environments.

In contrast to the previous models for multitask scenarios mainly focusing on the visual processes, this study develops quantitative models of the combined effects of auditory, tactile, and visual factors on multitasking performance. Specifically, the auditory and tactile modalities are modeled through not only the motor components (i.e., how to speak words; how to move body segments) but also the perception components (i.e., how to listen to words or tones).

1.5. Dissertation Structure

This dissertation includes nine chapters: this introductory chapter and eight additional chapters. Chapters 2 and 3 describe an experimental study and the corresponding modeling work regarding touchscreen gesture interactions. Chapter 2 introduces the experiment to quantify the performance of finger movement for the major touchscreen gestures (one-touch: dragging and swiping; two-touch: pinching and spreading), by investigating the effects of gesture direction and type on finger-touch input performance and subjective ratings. Chapter 3 describes the QN-based modeling developed for one of the tasks investigated in Chapter 2 (i.e., the dragging task), using three-dimensional motion tracking.

Chapters 4 and 5 describe an experimental study and the corresponding modeling work regarding a topic that combines audio/speech interaction and driving control. Chapter 4 describes the experiment to examine the effects of secondary-task modality and road curvature on driver behavior and performance (e.g., eye movement, subjective workload). Chapter 5 describes the QN-based modeling work of the tasks conducted in Chapter 4, to predict eyes-off-road behavior and workload in performing four different types of stimulus-response tasks (i.e., auditory-manual, auditory-speech, visual-manual, and visual-speech).

Chapters 6 and 7 introduce an experimental study and the corresponding modeling work regarding driving control. Chapter 6 describes the experiment to investigate the effects of road geometry (i.e., road curvature and curve direction) and existence of lead vehicle on driving performance. Chapter 7 introduces the QN-based modeling work to predict driver lateral control performance on curved roads, by integrating vehicle dynamics and reference trajectory tracking.

Chapter 8 introduces a real-world engineering application by describing the development and evaluation of a QN-based computational cognitive model for in-vehicle direct/indirect manual and speech interaction systems, which can be used as a usability testing tool of related products. Beyond the previous QN-MHP models using a simple voice-related operator (syllable-level), the model described in Chapter 8 is able to predict human performance for sentence-level voice commands. Also, visual feedback using colors and texts is implemented to model human indirect-manual controls.

Chapter 9 provides a summary of the results and conclusion from this dissertation study and ends with discussions for future research directions.

References

Barón, A., & Green, P. (2006). *Safety and usability of speech interfaces for in-vehicle tasks while driving: A brief literature review* (No. UMTRI-2006-5). University of Michigan, Transportation Research Institute.

Bi, L., Gan, G., Shang, J., & Liu, Y. (2012). Queuing network modeling of driver lateral control with or without a cognitive distraction task. *IEEE Transactions on Intelligent Transportation Systems*, *13*(4), 1810-1820.

Card, S., Moran, T. P., & Newell, A. (1983). The Psychology of Human-Computer Interaction. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cook, R. I., Potter, S. S., Woods, D. D., & McDonald, J. S. (1991). Evaluating the human engineering of microprocessor-controlled operating room devices. *Journal of Clinical Monitoring and Computing*, *7*(3), 217-226.

Feng, F., & Liu, Y. (2013). Computational modeling of feature and conjunction visual search tasks using Queuing Network-Model Human Processor (QN-MHP), *2nd International Digital Human Modeling Symposium Proceedings*.

Feyen, R., & Liu, Y. (2001, July). Modeling task performance using the queuing network-model human processor (QN-MHP). In *Proceedings of the 4th International Conference on Cognitive Modeling* (pp. 73-78).

Findlater, L., Froehlich, J. E., Fattal, K., Wobbrock, J. O., & Dastyar, T. (2013, April). Age-related differences in performance with touchscreens compared to traditional mouse input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 343-346). ACM.

Fuller, H. J., Reed, M. P., & Liu, Y. (2012). Integration of physical and cognitive human models to simulate driving with a secondary in-vehicle task. *IEEE Transactions on Intelligent Transportation Systems*, *13*(2), 967-972.

Gärtner, U., König, W., & Wittig, T. (2001). Evaluation of manual vs. speech input when using a driver information system in real traffic. In *Proceedings of the International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*.

Glass, J. (1999, December). Challenges for spoken dialogue systems. In *Proceedings of the 1999 IEEE ASRU Workshop*.

Hoggan, E., Nacenta, M., Kristensson, P. O., Williamson, J., Oulasvirta, A., & Lehtiö, A. (2013, October). Multi-touch pinch gestures: Performance and ergonomics. In *Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces* (pp. 219-222). ACM.

Jaimes, A., & Sebe, N. (2007). Multimodal human–computer interaction: A survey. *Computer Vision and Image Understanding*, *108*(1), 116-134.

Jennings, A., Ryser, S., & Drews, F. (2013, September). Touch screen devices and the effectiveness of user interface methods. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 57, No. 1, pp. 1648-1652). Sage CA: Los Angeles, CA: SAGE Publications.

John, B. E., & Kieras, D. E. (1996). The GOMS family of user interface analysis techniques: Comparison and contrast. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *3*(4), 320-351.

Jokinen, K., & Raike, A. (2003, September). Multimodality–technology, visions and demands for the future. In *Proceedings of the 1st Nordic Symposium on Multimodal Interfaces* (pp. 239-251).

Kieras, D. E. (1999). A guide to GOMS model usability evaluation using GOMSL and GLEAN3. *University of Michigan*, *313*.

Kobayashi, M., Hiyama, A., Miura, T., Asakawa, C., Hirose, M., & Ifukube, T. (2011). Elderly user evaluation of mobile touchscreen interactions. *Human-Computer Interaction–INTERACT 2011*, 83-99.

Kono, Y., Yano, T., & Sasajima, M. (1999). A generic framework for spoken dialogue systems and its application to a car navigation task. In *Intelligent Transportation Systems, 1999. Proceedings. 1999 IEEE/IEEJ/JSAI International Conference on*(pp. 728-733). IEEE.

Mann, W. C., Moore, J. A., & Levin, J. A. (1977, August). A comprehension model for human dialogue. In *Proceedings of the 5th international joint conference on Artificial intelligence-Volume 1* (pp. 77-87). Morgan Kaufmann Publishers Inc..

Lim, J. H., & Liu, Y. (2004, September). A Queuing Network Model for Visual Search and Menu Selection. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 48, No. 16, pp. 1846-1850). Sage CA: Los Angeles, CA: SAGE Publications.

Liu, Y., Feyen, R., & Tsimhoni, O. (2006). Queueing Network-Model Human Processor (QN-MHP). *ACM Transactions on Computer-Human Interaction*, *13*(1), 37–70.

Minsky, M. R. (1984, January). Manipulating simulated objects with real-world gestures using a force and position sensitive screen. In *ACM SIGGRAPH Computer Graphics* (Vol. 18, No. 3, pp. 195-203). ACM.

Poupyrev, I., & Maruyama, S. (2003, November). Tactile interfaces for small touch screens. In *Proceedings of the 16th annual ACM symposium on User interface software and technology* (pp. 217-220). ACM.

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

Shneiderman, B. (1991). Touch screens now offer compelling uses. *IEEE software*, *8*(2), 93-94.

Spence, C., & Driver, J. (1997). Cross-modal links in attention between audition, vision, and touch: Implications for interface design. *International Journal of Cognitive Ergonomics*.

Traum, D., Swartout, W., Gratch, J., & Marsella, S. (2008). A virtual human dialogue model for non-team interaction. *Recent trends in discourse and dialogue*, 45-67.

Tsimhoni, O., & Liu, Y. (2003, October). Modeling Steering Using the Queueing Network—Model Human Processor (QN-MHP). In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 47, No. 16, pp. 1875-1879). SAGE Publications.

Viviani, C. A. B., & Calil, S. J. (2015). Recommendation in the Use of Touchscreen Technology in Medical Devices. In *VI Latin American Congress on Biomedical Engineering CLAIB 2014, Paraná, Argentina 29, 30 & 31 October 2014* (pp. 754-756). Springer, Cham.

Wu, C., & Liu, Y. (2008). Queuing Network Modeling of Transcription Typing. *ACM Transactions on Computer-Human Interaction*, *15*(1), 1–45.

Zhao, G., & Wu, C. (2013). Mathematical modeling of driver speed control with individual differences. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *43*(5), 1091-1104.

Chapter 2.

An Experimental Investigation on the Effects of Touchscreen Gesture Type and Direction on Finger-touch Input Performance and Subjective Ratings

2.1. Introduction

To understand how humans use their fingers when operating touchscreen devices, it is important to examine the different characteristics of gestures performed on touchscreens. The corresponding research results can be used for designing touchscreen interfaces to better fit user needs. The characteristics of touchscreen gesture performance include finger movement time and speed, error rate, and finger trajectory for each touchscreen gesture. Over the past decades, touchscreen technologies have been developing rapidly and there have been many studies investigating the usability of touchscreen devices to understand user's needs (e.g., Sears, Revis, Swatski, Crittenden, & Shneiderman, 1993; Colle & Hiszem, 2004; Kim, Kwon, Heo, Lee, & Chung, 2014; Kim & Song, 2014; Garcia-Lopez, Garcia-Cabot, & de-Marcos, 2015). While the traditional touchscreen technologies allowed a limited number of finger-touch gestures (e.g., typing and clicking) with one finger (mostly an index finger), the type of touchscreen gestures and the fingers used have become more diverse recently.

The type of touchscreen finger gestures is generally composed of two categories: single-touch and multi-touch gestures. The single-touch gestures include clicking (or pressing, tapping), swiping (or flicking), and dragging (or scrolling, panning, sliding), whereas the multi-touch gestures include pinching (or zooming-out), spreading (or zooming-in). For the fingers used for

the single- and multi-touch gestures, Hinrichs and Carpendale (2011) found that the most common drag gestures were done by a single finger (48%), most often the index finger, in their observations of visitors using a tabletop touchscreen installed at a public exhibit. They also found that one of the common single-handed gestures for scaling the media items on the touchscreen device was the two-finger-pinch using the thumb and index fingers, and that the most frequently applied swipe action used the index finger.

In addition to the type of touchscreen gestures, it is necessary to pay attention to the directions of the gestures, because of their possible different functions and performance. For example, horizontal directions of drag and swipe gestures could mean forward or backward, whereas vertical directions for turning the sound volumes up and down (Burnett et al., 2013; Billinghurst & Vu, 2015). In addition, recent touchscreen technology allows the use of drag gestures to perform finger-writing or finger-painting, and the use of diverse directions of gesture for pattern lock instead of conventional password input (Andriotis et al., 2013). For pinch and spread gestures, different directions are used for more dedicated scaling in particular directions. Some recent studies have investigated a wide range of touchscreen gestures with respect to their different types and directions. For example, Gao and Sun (2015) studied the effects of age, type of touchscreen, inclination angle for clicking, dragging, zooming and rotating gestures. One of their findings was that right-handed subjects were more satisfied with rightward and downward drag gestures than leftward and upward dragging due to less movements of their upper arm. However, they did not investigate the direction effects of zooming and rotating in the study. Asakawa, Denneerlein, and Jindrich (2017) investigated task completion times for tapping, sliding, pinching, and spreading. They found that spreading was significantly slower than pinching ($p < .001$). In addition, two-finger gestures were slower than single-finger sliding (or

dragging). Specifically, pinch gestures took longer than sliding gestures regardless of their directions (i.e., up, down, left, and right), while spread gestures were significantly slower than sliding gestures only in three directions (i.e., up, down, and left) except in the right direction. Hoggan et al. (2013) studied performance (i.e., task completion time and error rate) of pinch and spread gestures. They found that the task completion time of pinch gesture was shorter in the horizontal direction than the diagonal direction (i.e., rotating 22.5 degrees counter clockwise from the horizontal line). Although the previous studies are important in exploring the diverse touchscreen gesture types in multiple directions, the range of gesture directions they employed were limited and it is still not known whether the effect of gesture directions on human performance varies by the gesture type.

In this study, I aimed to quantify the performance of finger movement for the major touchscreen gestures, by investigating the effects of gesture direction and type on finger-touch input performance and subjective ratings. Specifically, task completion time, task failure status, error rate, and subjective ratings (i.e., performance and physical demand) were measured with respect to the one-touch (i.e., dragging and swiping) and the two-touch (i.e., pinching and spreading) touchscreen gestures. In the experiment, using a touchscreen tablet, participants were asked to use their index fingers for the one-touch gestures, and their thumb and index fingers for the two-touch gestures.

2.2. Method

2.2.1. Participants

Twenty right-handed students (11 males and 9 females) ranging in age from 18 to 30 (M = 23.8, SD = 3.3) were recruited from a university. They had experience with touchscreen

devices for more than 6 years on average. They self-reported to have normal or corrected-to-normal vision and none of them had physical discomfort and impairments. The experiment took approximately one hour and they were paid for their participation.

2.2.2. Apparatus

A 9.7-inch touchscreen device (iPad; 1024 × 768; Multi-Touch display) was used and it was mounted on a 95-cm high table, with the display surface facing the ceiling (i.e., no inclination angle). For this experiment, a web-based interface software was developed and connected through the wireless Internet for data transfer. The software was programmed to automatically record the finger-touch data, including two-dimensional coordinates and timestamps, by two categories: touch-in and touch-out. In this study, touch-in refers to an action when a finger arrives from the air to the display surface, whereas touch-out refers to an action when a finger leaves from the display surface to the air. Whenever the finger touched in or out on the display surface, a 20-pixel-radius black circle was shown on the display, as a visual feedback, to inform participants that the finger-touch was recognized and recorded in the interface system in real time.

2.2.3. Finger-touch gesture task

2.2.3.1. One-touch gesture task

For the one-touch gesture task, 8 directions covering a circular two-dimensional region were used: 0, 45, 90, 135, 180, 225, 270, and 315 degrees of angular direction, based upon the Cartesian coordinates systems, as shown in Figure 2.1-(a). These 8 directions were used for both the drag and swipe one-touch gesture tasks. Participants were asked to place their right index finger on a green circle with a hole (i.e., a center circle or a hollow circle) and to move their index finger toward each circle without a hole (i.e., a target circle or a filled circle). The distance

between the center and the target circles (40 mm) was determined by a pilot test using five participants, which found an average dragging distance on the touchscreen device, allowing them to reach to all the eight directions with the same distance. The size of the two circles was fixed during the experiment and the radius of all the circles was 20 pixel or 9 mm. A successful one-touch gesture task was defined as the situation when the Euclidean distance between the centers of the black (i.e., the visual feedback circle) and green circles was equal to or less than 20 pixels (also called a match of the black and green circles). For the dragging task, participants were asked to draw a line from a hollow circle to the other circle in a normal pace. They were asked to accomplish the dragging task with two successes on the initial and final matches (i.e., center and target-circle matches). For the swiping task, participants were asked to flick their index finger from the center circle toward the direction of the target circle. They were asked to accomplish the swiping task successfully only once on the initial match (i.e., center-circle match) and swipe toward the direction of the target circle.

2.2.3.2. Two-touch gesture task

For the two-touch gesture task, 4 directions were selected (see Figure 2.1-(b), whose A-D directions correspond to the 0-180, 45-225, 90-270, 135-315 degrees in the one-touch gesture task), which are the most often used directions for the two-touch gesture with relatively less discomfort than the other unselected directions, such as 180-0 (in which the locations of thumb and index fingers are reversely assigned, compared to the 0-180 direction). Participants were asked to place their right thumb on a red circle with a hole (i.e., a red hollow circle for thumb) and index finger on a green circle with a hole (i.e., a green hollow circle for index finger) and to move the two fingers (thumb and index) to the same-colored circles without a hole (i.e., target filled circles for both the thumb and the index fingers). The distance between the hollow circles

was 80 mm, while the distance between the filled circles was 20 mm. Like the one-touch gesture task, the size of the four circles was fixed during the experiment and the radius of all circles was 20 pixel or 9 mm. Like the one-touch task, a successful two-touch task was defined as the situation when the Euclidean distance between the centers of the black and green/red circles was equal to or less than 20 pixels for both fingers. For the pinch and spread tasks, participants were asked to draw lines from hollow circles to filled circles in a normal pace toward inside (for pinching) and outside (for spreading). Participants were asked to accomplish the pinching and spreading tasks with four successes on initial (using two fingers) and final matches (using the same two fingers).

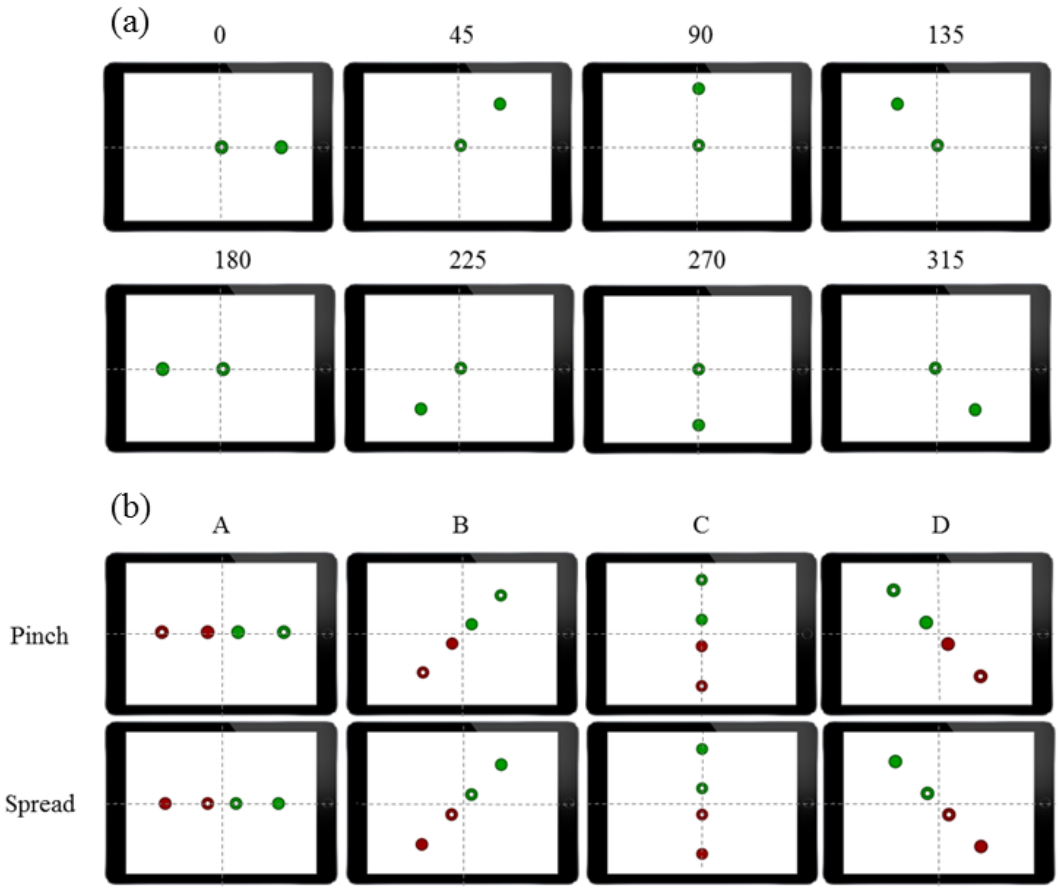


Figure 2.1 – Directions for the touch gesture tasks

(a) Eight angular directions (unit: degree) for the one-touch gesture task (moving from hollow to filled circles), (b) Four directions (*A, B, C, and D*) for the two-touch gesture task (green for index finger, red for thumb; moving from hollow to filled circles)

2.2.4. Design of Experiments

The two independent variables were gesture direction and gesture type. Gesture direction had 8 levels (i.e., 0, 45, 90, 135, 180, 225, 270, and 315 degrees) in the one-touch gesture task, and 4 levels (i.e., A, B, C, and D) in the two-touch gesture task. Gesture type had 2 levels for both one-touch (i.e., drag and swipe) and two-touch (i.e., pinch and spread) gesture tasks. Three repetitions of each level were performed by the participants. Twenty participants performed 6 sessions (within-subject): 4 sessions for the one-touch task and 2 sessions for the two-touch task. Two of the sessions for the one-touch task were the dragging task (one for horizontal-vertical directions, another for diagonal directions) and another 2 sessions for the one-touch task were the swiping task (one for the horizontal-vertical directions, another for the diagonal directions). The 2 sessions for the two-touch task were pinching and spreading, respectively. The order of the six sessions and each level for the two independent variables were randomized and counterbalanced among the participants.

2.2.5. Dependent variables

2.2.5.1. Task completion time

Many touchscreen gesture studies used gesture duration (or task completion time) as a measurement of gesture performance. For example, Hoggan et al. (2013) measured the time between the movement onset and the removal of both fingers from the display surface as the pinch/spread gesture duration. Gao and Sun (2015) measured the amount of time that participants spent in completing the task after they clicked the “start” button. In this study, the completion times of the one-touch gesture task were measured as the time gap between the touch-in and

touch-out moments, only for the success trials. The completion times of the two-touch gesture task were measured as the time gap between the first touch-in moment of the two fingers and the second touch-out moment of the two fingers, only for the success trials.

2.2.5.2. Task failure

All the touchscreen gesture tasks used in this study required not only an exact touching action but also an exact processing/moving action for their success. In other words, it is important to confirm whether the touchscreen device detects the user's input command, thereby identifying the positions of the touch-in and touch-out on the display surface. In this regard, the task success in this study was defined as whether both touch-in and touch-out actions were performed correctly, meaning the center of black circles appeared due to both touch-in and touch-out actions was within 20 pixels from the center of green/red circles. I believe a task failure causes incorrect input message or information to the device. Thus, a task failure was also used as a dependent measure, referring to either touch-in or touch-out as a failure.

2.2.5.3. Error rate

Since the one-touch gestures have two moments of finger-touch on the display (i.e., one touch-in and one touch-out), it is relatively easy to investigate whether or not the task was a failure. In contrast, since the two-touch gestures have four moments of touch (i.e., two touch-ins and two touch-outs), a more precise measure was needed to examine the two-touch gesture failures. Thus, as an additional analysis for the two-touch gesture, the error rate was used, which refers to the percentage of all the four failure possibilities.

2.2.5.4. Subjective ratings

Subjective ratings were measured by two categories using a scale ranging from 1 (lowest) to 7 (highest): performance (i.e., how successful you were in accomplishing; i.e., shorter task

completion time and lower error rate) and physical demand (i.e., how physically demanding the task was). The performance and physical demands were rated on a seven-point scale for three directions, including horizontal, vertical, and diagonal, as well as for two gesture types (i.e., dragging and swiping for the one-touch task; pinching and spreading for the two-touch task). The reason why the subjective performance was measured in this study was to compare it with the objective measures (i.e., was there consistency between what participants perceived and what they actually did) and to understand which condition makes participants feel physically demanding, as a non-intrusive measure.

2.2.6. Procedure

Prior to the experiment, all the participants were given an informed consent form reviewed by the university's institutional review board, and provided with a brief introduction of the purposes and nature of the study. Participants were instructed to stand at the workspace (the center of the device was located at the right-front side approximately 30 cm away from the right edge of participant's torso) and adjust the height of the touchscreen device for their most natural and comfortable posture (the average height of the touchscreen device was approximately 105 cm from the floor). After the participants practiced for approximately 10 - 15 minutes, they were asked to perform 6 sessions for the two main tasks: 4 sessions for the one-touch task and 2 sessions for the two-touch task. Participants were informed that the goal of the experiment is to achieve a task completion time as short as possible and an error rate as low as possible for each task. After completing each session, participants were asked to rate their subjective performance and physical demand. At the end of all the six sessions, which lasted for approximately an hour, participants signed a payment form.

2.2.7. Data analysis

The dependent measures of task completion time, error rate, and subjective ratings were analyzed by repeated-measures two-way ANOVA (8 directions × 2 gesture type for the one-touch task, whereas 4 directions × 2 gesture type for the two-touch task) using SPSS Statistics 24. Prior to performing the ANOVA tests, the three dependent variables were verified whether they violate the normality assumption of ANOVA, using the criteria of skewness (< 2) and kurtosis (< 7) (Curran, West, and Finch, 1996). Additionally, the Levene's test for homogeneity of variance was used to investigate the assumption of homogeneity of variance across groups. Effect sizes were measured by a partial eta-squared (η_p^2). To determine whether there is an association between the task failure status and the categorical independent variables (i.e., gesture direction and type), the Chi-square test was conducted and Cramer's V was used as a measure of its effect sizes. To investigate the effect of horizontal-vertical-diagonal direction on the dependent variables, the gesture directions were also categorized by three directions: horizontal (0 and 180 degrees in the one-touch task; A in the two-touch task), vertical (90 and 270 degrees; C), and diagonal (45, 135, 215, and 275 degrees; B and D). The significance level for all statistical tests was $p \leq .05$. The adjusted standardized residuals were used as a post-hoc criterion (> 1.96) in the Chi-square test (García-pérez & Núñez-antón, 2003).

2.3. Results

Tables 2.1 and 2.2 provide a summary of statistical effects for each task, found from the two-way ANOVA tests. The results of chi-square tests for each task were presented in Tables 2.3 and 2.4 To investigate the effect of directions in horizontal (H), vertical (V), and diagonal (D)

trajectories from the participants' view, a new variable (called HVD in the tables) was created and analyzed in SPSS as well.

Table 2.1 – ANOVA summary of main effect on one-touch gesture measures

	Factors	*MSE*	*F*	*df1*	*df2*	*p*
Task completion time (ms)	Gesture type	222184369.5	1403.062	1	892	< .001**
	Direction (8 levels)	154437.308	1.393	7	892	.205
	Gesture type	193588385.6	1218.127	1	902	< .001**
	Direction (HVD)	254477.308	.801	2	902	.449
Subjective performance	Gesture type	1.2	.861	1	114	.335
	Direction (HVD)	6.3	4.52	2	114	.013*
Subjective physical demand	Gesture type	.533	.205	1	114	.652
	Direction (HVD)	16.233	6.227	2	114	.003**

*p<.05, **p<.01

Table 2.2 – ANOVA summary of main effect on two-touch gesture measures

	Factors	*MSE*	*F*	*df1*	*df2*	*p*
Task completion time (ms)	Gesture type	1134060.4	2.357	1	149	.127
	Direction (4 levels)	1408826.4	2.928	3	149	.036*
	Gesture type	843729.0	1.684	1	151	.196
	Direction (HVD)	517209.0	1.032	2	151	.359
Error rate (%)	Gesture type	59.6	10.187	1	443	.002**
	Direction (4 levels)	20.3	3.46	3	443	.016*
	Gesture type	49.3	8.357	1	445	.004**
	Direction (HVD)	15.7	2.66	2	445	.071
Subjective performance	Gesture type	1.406	.801	1	152	.372
	Direction (4 levels)	23.656	13.48	3	152	< .001**
Subjective physical demand	Gesture type	.756	.363	1	152	.548
	Direction (4 levels)	42.973	20.602	3	152	< .001**

*p<.05, **p<.01

Table 2.3 – One-touch task's success/failure status by gesture direction and type

	Gesture type (2 levels): frequency (%)								χ^2	*df*	*p*
	Drag	Swipe							6.179	1	.013*
Task success	446 (93.1%)	462 (96.7%)									
Task failure	33 (6.9%)	16 (3.3%)									
	Direction (8 levels - degree): frequency (%)								χ^2	*df*	*p*
	0	45	90	135	180	225	270	315	17.96	7	.012*
Task success	116 (96.7%)	118 (99.2%)	117 (97.5%)	112 (93.3%)	111 (92.5%)	110 (92.4%)	116 (97.5%)	108 (90.0%)			
Task failure	4 (3.3%)	1 (.8%)	3 (2.5%)	8 (6.7%)	9 (7.5%)	9 (7.6%)	3 (2.5%)	12 (10.0%)			
	Direction (HVD): frequency (%)								χ^2	*df*	*p*
	Horizontal	Vertical	Diagonal						4.709	2	.095
Task success	227 (94.6%)	233 (97.5%)	448 (93.7%)								
Task failure	13 (5.4%)	6 (2.5%)	30 (6.3%)								

*p<.05, **p<.01

Table 2.4 – Two-touch task's success/failure status by gesture direction and type

	Gesture type (2 levels): frequency (%)				χ^2	*df*	*p*
	Pinch	Spread			6.179	1	.013*
Task success	71 (29.6%)	86 (40.8%)					
Task failure	169 (70.4%)	125 (59.2%)					
	Gesture direction (4 levels): frequency (%)				χ^2	*df*	*p*
	A	B	C	D	12.522	3	.006**
Task success	25 (22.3%)	49 (43.4%)	44 (39.6%)	39 (33.9%)			
Task failure	87 (77.7%)	64 (56.6%)	67 (60.4%)	76 (66.1%)			
	Gesture direction (HVD): frequency (%)				χ^2	*df*	*p*
	Horizontal	Vertical	Diagonal		10.279	2	.006**
Task success	25 (22.3%)	44 (39.6%)	88 (38.6%)				
Task failure	87 (77.7%)	67 (60.4%)	140 (61.4%)				

*p<.05, **p<.01

2.3.1. One-touch gesture task

2.3.1.1. Touch point distribution

Figures 2.2-(a) and 2.2-(b) show the distribution of the one-finger touch locations. The red and black marks indicate the location of touch-in and touch-out, respectively. Most of the touch-out points (i.e., the black marks) for the dragging gesture were located in the eight target circles. However, the touch-out points during swiping were on the way toward the target circles, but not on the circles.

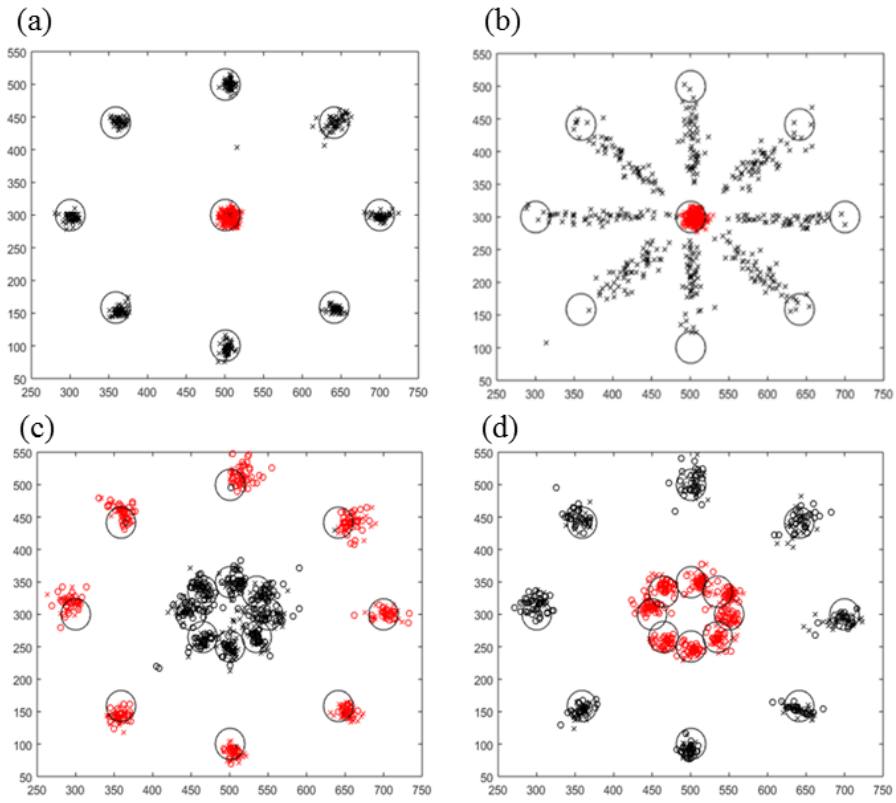


Figure 2.2 – Distribution of one-finger touch points ((a) dragging and (b) swiping) and two-finger touch points ((c) pinching and (d) spreading (×- first touch, ○- second touch)) on the display

2.3.1.2. Task completion time

The effect of gesture type was significant: F (1,892) = 1403.1, $p < .001$, $\eta_p^2 = .611$. As shown in Figure 2.3-(a), the completion time of the drag gesture (mean ($\bar{x}$) = 1272.2, standard error of the mean (*SEM*) = 26.4) was significantly longer than that of the swipe gesture ($\bar{x}$ = 285.1, *SEM* = 5.3). However, no significant differences were found among the eight angular directions and the three categorical directions (i.e., HVD).

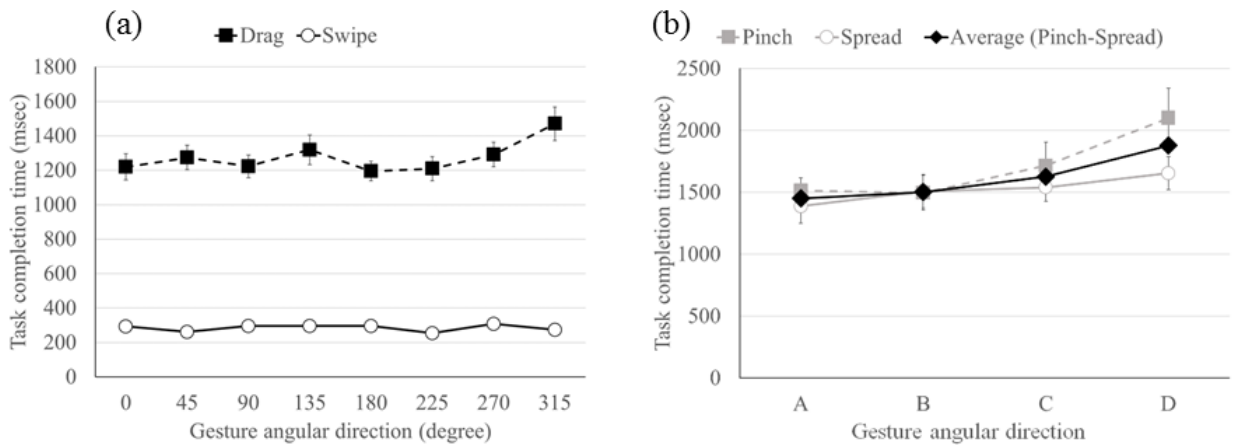


Figure 2.3 – Task completion time of (a) one-touch and (b) two-touch gestures by gesture angular direction (Error bars = *SEM*)

2.3.1.3. Task failure

The result of Pearson Chi-square test showed a significant relationship between task failure status and the gesture type, χ^2 (1, N = 957) = 6.179, p = .013, but with weak strength of the relationship (Cramer's V = .08). As shown in Figure 2.4-(a), the drag gesture task (6.9 %) had higher percentage of task failure than the swiping task (3.3 %). For the relationship between the task failure and the angular direction, there was also a significant effect of the directions on the task failure (χ^2 (7, N = 957) = 17.960, p = .012) with a weak relationship (Cramer's V = .137). In particular, task failure occurred most often at 315 degrees (10.0 %, adjusted standardized residual = 2.4), shown in Figure 2.4-(b).

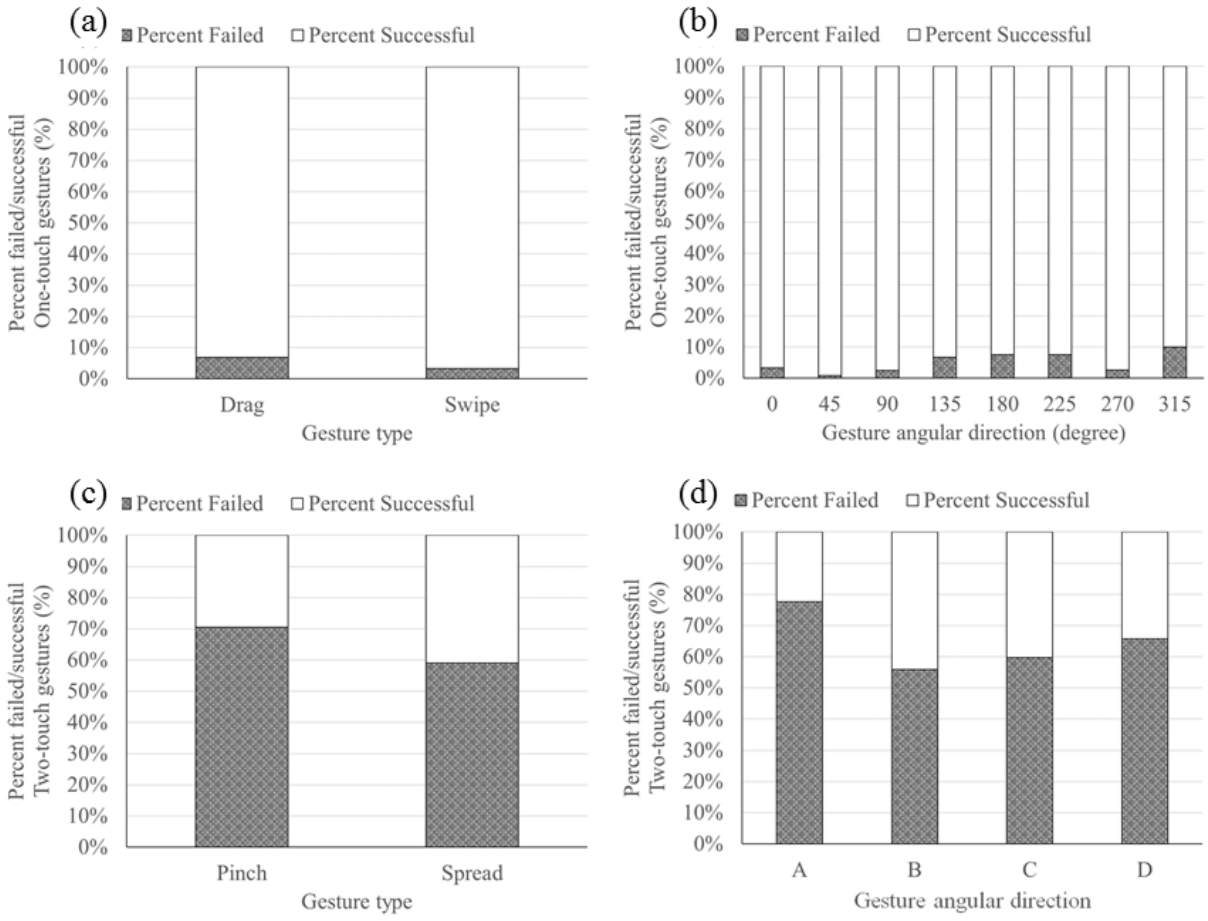


Figure 2.4 – Percentage of (a-b) one-touch and (c-d) two-touch task's failure/success by gesture type and gesture angular direction

2.3.1.4. Subjective ratings

As shown in Figure 2.5-(a), ANOVA for the performance revealed a significant main effect of the gesture direction (F (2,114) = 4.52, p = .013, η_p^2 = .073), but no significant main effect of the gesture type nor any interaction effect. Pairwise comparisons, using a Tukey HSD procedure, showed that the performance when participants moved their finger along the horizontal direction ($\bar{x}$ = 5.6, SEM = .187) was significantly higher (p = .015) than the diagonal direction ($\bar{x}$ = 4.9, SEM = .187). ANOVA for the physical demand showed a statistically significant difference on the gesture direction (F (2,114) = 6.23, p = .003, η_p^2 = .098), but no

significant main effect of the gesture type nor interaction between two factors. Post hoc test revealed that physical demands during diagonal ($\bar{x}$ = 3.4, *SEM* = .255) and vertical movements ($\bar{x}$ = 3.3, *SEM* = .255) were higher than horizontal movement ($\bar{x}$ = 2.2, *SEM* = .255).

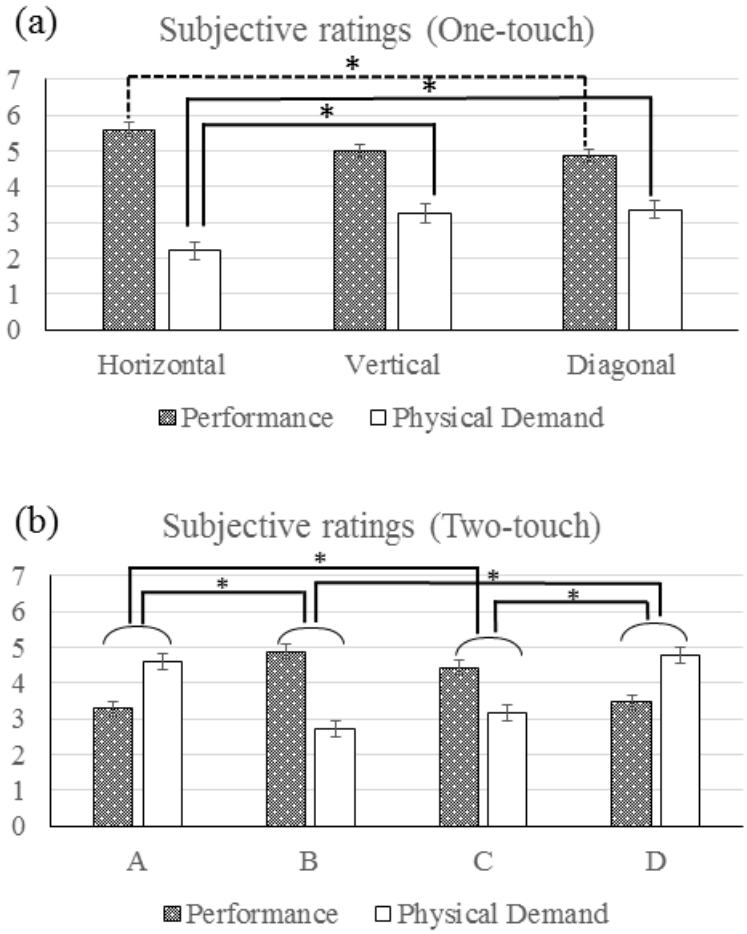


Figure 2.5 – The main effect of (a) one-touch and (b) two-touch gesture's direction on subjective ratings (Error bars = *SEM*)

2.3.2. Two-touch gesture task

2.3.2.1. Touch point distribution

Figures 2.2-(c) and 2.2-(d) show the distribution of the two-finger touch locations. As also indicated in Figures 2.2-(a) and 2.2-(b), the red and black marks indicate the touch-in and touch-out points on the display, respectively. Additionally, ×-marks indicate the locations of the

fingertip's first touch, whereas ○-marks indicate the locations of the fingertip's second touch between the thumb and index fingers. Of the two fingers, the index finger was touched-in earlier and touched-out earlier than the thumb (see Table 2.5). To estimate the width of zooming-in and out gestures, participants tended to use the index finger to reach the display first. It might be easier to touch the display surface with the index finger than the thumb, because the index finger is longer than the thumb.

Table 2.5 – Percentage of index finger's touch-in/out as a first touch (%)

		A	B	C	D
Pinch	Touch-in	62.1	65.6	69.5	61.3
	Touch-out	64.2	67.3	59.3	59.9
Spread	Touch-in	70.4	61.5	53.8	52.8
	Touch-out	63.4	63.6	65.6	67.8

2.3.2.2. Task completion time

An ANOVA performed on task completion time revealed a significant effect of gesture direction ($F(3,149) = 2.928$, $p = .036$, $\eta_p^2 = .056$), but no effect of gesture type nor interaction effect. As shown in Figure 2.3-(b), the mean task completion time of the two-touch gesture was the longest in the D direction ($\bar{x} = 1885.6$, $SEM = 83.4$), whereas it was the shortest in the A direction ($\bar{x} = 1453.9$, $SEM = 137.7$).

2.3.2.3. Task failure

A Chi-square test revealed a mildly significant relationship between the task failure status and the gesture type: $\chi^2 (1, N = 451) = 6.179$, $p = .013$, Cramer's V = .117. As shown in Figure 2.4-(c), pinch gesture (70.4 %) had more task failure than spread gesture (59.2%). There was also a mildly significant relationship between the task failure status and the gesture direction: $\chi^2 (3, N = 451) = 12.522$, $p = .006$, Cramer's V = .006. As shown in Figure 2.4-(d), the two-touch gesture

had the most task failure in the A direction (77.7 %) and the least task failure in the B direction (56.6%). In addition, a mildly significant relationship between the task failure status and the gesture direction (HVD) was found (χ^2 (2, N = 451) = 10.279, p = .006, Cramer's V = .151): the most task failure occurred in the horizontal direction (77.7%), whereas the least task failure occurred in the vertical direction (60.4%).

2.3.2.4. Error rate

Error rate for the two-touch gesture task was significantly affected by the gesture type (F (1,443) = 10.187, p = .002, η_p^2 = .022), and the gesture direction (F (3,443) = 3.460, p = .016, η_p^2 = .023), but no interaction effect was found. For the comparison between gesture types, shown in Figure 2.6-(a), error rate ($\bar{x}$ = 29.5%, *SEM* = 1.7) was significantly higher for pinch gesture than spread gesture ($\bar{x}$ = 22.3%, *SEM* = 1.6). Post-hoc Tukey's comparisons showed that the error rate for the A direction ($\bar{x}$ = 30.4%, *SEM* = 2.2) was significantly higher than the one for the B direction ($\bar{x}$ = 21.5%, *SEM* = 2.2; p = .031), but other significant differences among the four directions were not found (see Figure 2.6-(b)).

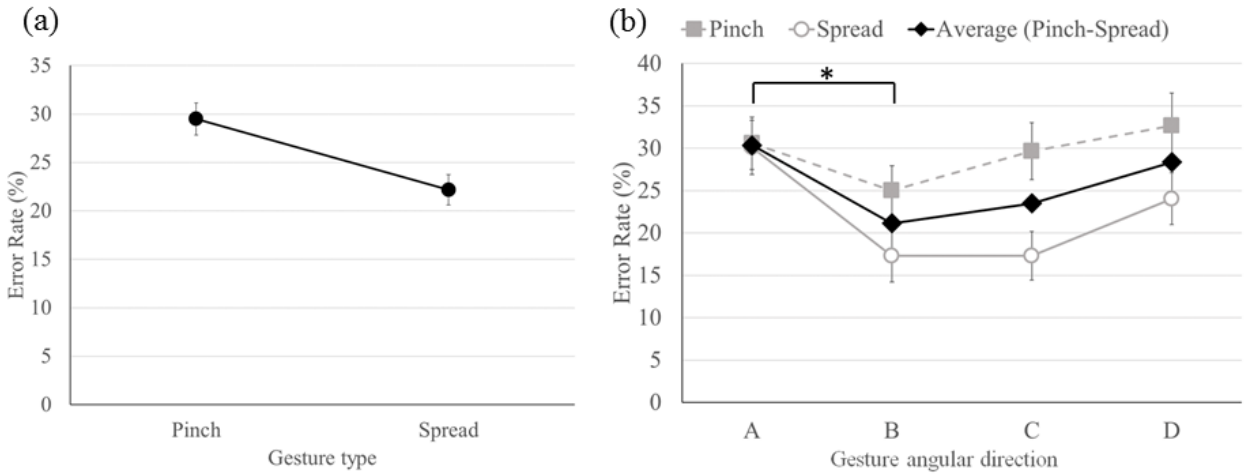


Figure 2.6 – Error rate of two-touch gestures by (a) gesture type and (b) gesture direction (Error bars = *SEM*)

2.3.2.5. Subjective ratings

The significant effect of gesture direction (F (3, 152) = 13.48, $p < .001$, $\eta_p^2 = .21$) showed the highest performance for the B direction ($\bar{x} = 4.9$, $SEM = .209$), followed by the C direction ($\bar{x} = 4.4$, $SEM = .209$), the D direction ($\bar{x} = 3.5$, $SEM = .209$), and the A direction ($\bar{x} = 3.3$, $SEM = .209$), seen in Figure 2.5-(b). Post-hoc Tukey's tests showed that the mean performance for each direction differed significantly from each other direction (all $p < .05$), except between A and D, and between B and C directions. The physical demand that participants reported was significantly affected by direction (F (3, 152) = 20.6, $p < .001$, $\eta_p^2 = .289$), in the sequence of D direction ($\bar{x} = 4.8$, $SEM = .228$) > A direction ($\bar{x} = 4.6$, $SEM = .228$) > C direction ($\bar{x} = 3.2$, $SEM = .228$) > B direction ($\bar{x} = 2.7$, $SEM = .228$), but no effect of gesture type and interaction effect were found. Similar to the subjective performance results, Tukey's HSD post-hoc test revealed that mean physical demand for each direction differed significantly from other directions (all ps < .001), except two comparisons (A-D and B-C).

2.3.3. Result Summary

Table 2.6 summarizes the differences in dependent measures by gesture type and direction for both one-touch and two-touch gestures. Note that this table only contains the results that are statistically significant.

Table 2.6 – Result summary

Measures (unit)	One-touch	Two-touch
Task completion time (msec)	[Type] • Drag (1272.2) > Swipe (285.1)	[Direction] • D (1885.6; longest), A (1453.9; shortest)
Task failure (%)	[Type] • Drag (6.9) > Swipe (3.3) [Direction] • 315 degree (10.0; highest), 45 degree (.8; lowest)	[Type] • Pinch (70.4) > Spread (59.2) [Direction] • A (77.7; highest), B (56.6; lowest) • Horizontal (77.7; highest), Vertical (60.4; lowest)
Error rate (%)	N/A	[Type] • Pinch (29.5) > Spread (22.3) [Direction] • A (30.4; highest), B (21.5; lowest)
Subjective performance (1-7)	[Direction] • Horizontal (5.6; highest), Diagonal (4.9; lowest)	[Direction] • B (4.9; highest), A (3.3; lowest)
Subjective physical demand (1-7)	[Direction] • Diagonal (3.4; highest), Horizontal (2.2; lowest)	[Direction] • D (4.8; highest), B (2.7; lowest)

2.4. Discussion

2.4.1. Effects of touchscreen gesture type

The ANOVA test revealed a significant effect of the gesture type on task completion time for the one-touch gesture task: swiping was approximately 4.5 times faster than dragging. This was one of the expected results that swiping is faster than dragging, but how much faster swiping is than dragging was not measured in the previous studies (e.g., Wolf, Schleicher, & Rohs, 2014). In contrast, there was no significant effect of the gesture type on task completion time for

the two-touch gesture task. Interestingly, however, this finding differs from the results of some previous studies. Findlater et al. (2013) reported that pinching took a longer time than spreading ($p < .002$), without considering the direction of the two-touch gestures. Hoggan et al. (2013) revealed that spreading took a longer time than pinching ($p < .05$), considering only the directions between A and B defined in the current study. In this regard, whether there is a significant difference of task completion times between pinching and spreading might vary depending on factors other than direction, such as the distance between the two finger-touches on the display surface, size of touchscreen devices, or participant's anthropometric data (e.g., finger and hand sizes).

With respect to the task failure status, a significant relationship was found with the gesture types, for both the one-touch and two-touch gesture tasks. In the one-touch gesture task, dragging (6.9%) caused more task failures than swiping (3.3%). There were probably more chances to fail during dragging than swiping, because participants had to pay attention to both the initial and final targets for drag but only the initial target for swipe. In the two-touch gesture task, the percentage of task failures also differed by the gesture type: pinching (70.4%) and spreading (59.2%). Similarly, also for the error rate of two-touch gestures, there was a significant effect of the gesture type: pinch had 1.3 times higher error rate than spread. This result indicated a higher possibility to fail to touch-in and touch-out on/from the two target circles with pinching than spreading. Participants were a bit more likely to succeed in spreading which required relatively easier touch on the final targets separated with enough distance, compared to pinching. Interestingly, in contrast to these objective measures collected by the system, there was no difference in subjective ratings between the different gesture types. The performance rated by the

participants did not differ by the gesture types because the tasks were relatively easy and there was no large difference in the physical demands between gesture types.

2.4.2. Effects of touchscreen gesture direction

The effect of gesture direction on the task completion time was found only for the two-touch gesture task: when the index finger was located on the horizontal line (i.e., A direction), it took the shortest time for the pinching and spreading. The finding in this study is consistent with Hoggan et al. (2013), reporting that pinch/spread in smaller angles from the horizontal line (i.e., getting from D to A directions) were performed in shorter duration. With respect to the task failure status and error rate, compared to the task completion time, more clear effects of gesture direction were revealed. There was a significant relationship for task failure status with gesture direction, for both the one-touch and two-touch gesture tasks. For the one-touch gesture task, the highest rate of task failures occurred in the direction of 315 degrees (24.5% among all directions) and it was probably due to the occlusion by participants' right hand. Several participants mentioned that it was hard to find the target circle located in the right-bottom area in the display when they performed one-touch gesture task. For the two-touch gesture task, the highest rate of task failures occurred in the A direction (29.6% of among all directions), followed by the D direction (25.6%).

On the other hand, error rate for the two-touch gesture task was significantly affected by the gesture direction, with the highest rate in the A direction (30.4%) and the lowest in the B direction (21.5%). For the subjective ratings, significant effects of the gesture direction were found. Participants reported that they felt one-touch gestures in the horizontal directions resulted in better performance and lower physical demand, than in vertical and diagonal directions. For the two-touch gestures, participants tended to perceive better performance and lower physical

demand when they performed the tasks in the B and C directions, and worse performance and higher physical demand in the A and D directions. This subjective performance result is in agreement with the task failure and error rate results.

In summary, for the two-touch gestures, the A and D directions probably caused postures that are physically more demanding. That is, participants might need to twist their shoulders, upper/lower arms, or wrists more to perform the pinching and spreading tasks for the A and D directions, than the B and C directions. Thus, the possibility of task failure and error occurrence might be higher for the A and D directions. However, participants were able to put their right upper arms to the right torsos while performing the tasks so that they can have stable postures in the A direction, than the other three directions. For this reason, the task completion time at the A direction might be the shortest, once they succeeded in performing the tasks.

2.4.3. Limitations

There are several limitations of this study that should be taken into account when interpreting the results. First, it should be noted that all the results of this study were obtained from fixed and static positions of touchscreen displays. I measured the touchscreen gesture performance on the display surface in fixed positions, after I asked participants to self-adjust the height of the touchscreen device for their most natural and comfortable posture. The reason why a fixed and static setting was used in this study was to make the experimental setting purely static without considering any other factors, such as inclination angle, use of left hand, and so on. The workplace might not be a user-friendly setting for all jobs, but relevant for some jobs with standing postures such as a receptionist standing behind a desk. Future studies could conduct the usability test of touchscreen device in more dynamic conditions, such as holding the device with one hand and performing the gesture task with another hand. Second, the physical demand was

measured only with a subjective rating, an inexpensive and a non-intrusive method. Although the participants were asked to report their rating right after each session so that they could answer the questions clearly without loss of memory or perception of what they were performing for each session, more objective and systematic methods, such as electromyography, could also be explored (e.g., Kietrys et al., 2015). Third, it was not considered that the errors occurred by different level of touchscreen's sensitivity. Each device has different level of touching sensitivity. For example, some recent touchscreen devices have a function of proximity sensing, indicating the touching with high sensitivity by detecting a simple placement of a hand near the screen (Barrett & Omote, 2010). Thus, future studies could use multiple devices with different levels of sensitivity to avoid the errors caused by the device's own properties.

2.5. Implications and future studies

This study provides quantitative comparisons of human performance and subjective workload of different gesture types and directions for most often used one-touch and two-touch touchscreen gestures. The findings in this study can contribute to future touchscreen interface designs. First, task completion time is associated with the time that the participants' fingers touched on the touchscreen surface. For example, the result that dragging took 4.5 times longer than swiping could be used when designing buttons or spaces on the touchscreen for one-touch gestures: proportionally 4.5 time more space is required for dragging than swiping. In addition, based upon the result that two-touch gestures in the D direction took longer time than the A direction, more space on the touchscreen would be required for tasks in the D direction than the A direction. Secondly, the results of task failure and error rate will be useful for touchscreen interface designs, since they are related to whether the touch-input commands are correctly

applied to the touchscreen device. For example, swiping and spreading would be better to avoid input errors than dragging and pinching, respectively. With respect to the direction of gesture, the two-touch gesture in the A direction should be avoided because it has higher input error rates. Also, another example of design recommendation is to induce users not to perform the drag and swipe gestures in the direction of 315 degrees, because it is associated with the highest task failure. Lastly, the results of subjective physical demand can be used for user-friendly touchscreen designs. For example, it is recommended to the interface designers to allow the users to perform the pinching and spreading in the B or C direction, rather than A or D direction, by assigning the corresponding buttons or interface on the display, in order to reduce the physical demand.

Future research plans include developing a computational model to predict the task completion time and workload for both one-touch and two-touch gesture tasks. I have developed a queuing theory-based predictive model that estimates the touchscreen swipe gesture's execution time (Jeong and Liu, 2016). As an extension of this modeling study for swipe gesture, I plan to develop computational models for other major touchscreen gestures. According to the literature, finger size is one of the factors that may affect human performance of touchscreen gesture tasks (e.g., Scott & Conzola, 1997; Bergstrom-Lehtovirta, & Oulasvirta, 2014). In this regard, user anthropometric data (e.g., hand and finger sizes) will be used as predictors of the models. I expect these predictive models will help interface designers to find usability solutions at an early stage of system development, thereby reducing labor and time cost of usability evaluation.

References

Andriotis, P., Tryfonas, T., Oikonomou, G., & Yildiz, C. (2013, April). A pilot study on the security of pattern screen-lock methods and soft side channel attacks. In *Proceedings of the sixth ACM conference on Security and Privacy in Wireless and Mobile Networks* (pp. 1-6). ACM.

Asakawa, D. S., Dennerlein, J. T., & Jindrich, D. L. (2017). Index finger and thumb kinematics and performance measurements for common touchscreen gestures. *Applied Ergonomics*, *58*, 176-181.

Barrett, G., & Omote, R. (2010). Projected-capacitive touch technology. *Information Display*, *26*(3), 16-21.

Bergstrom-Lehtovirta, J., & Oulasvirta, A. (2014, April). Modeling the functional area of the thumb on mobile touchscreen surfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1991-2000). ACM.

Billinghurst, S. S., & Vu, K. P. L. (2015). Touch screen gestures for web browsing tasks. *Computers in Human Behavior*, *53*, 71-81.

Burnett, G., Crundall, E., Large, D., Lawson, G., & Skrypchuk, L. (2013, October). A study of unidirectional swipe gestures on in-vehicle touch screens. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 22-29). ACM.

Colle, H. A., & Hiszem, K. J. (2004). Standing at a kiosk: Effects of key size and spacing on touch screen numeric keypad performance and user preference. *Ergonomics*, *47*(13), 1406-1423.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*(1), 16.

Findlater, L., Froehlich, J. E., Fattal, K., Wobbrock, J. O., & Dastyar, T. (2013, April). Age-related differences in performance with touchscreens compared to traditional mouse input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 343-346). ACM.

Gao, Q., & Sun, Q. (2015). Examining the usability of touch screen gestures for older and younger adults. *Human factors: The Journal of the Human Factors and Ergonomics Society*, *57*(5), 835-863.

Garcia-Lopez, E., Garcia-Cabot, A., & de-Marcos, L. (2015). An experiment with content distribution methods in touchscreen mobile devices. *Applied Ergonomics*, *50*, 79-86.

Garcia-Perez, M. A., & Nunez-Anton, V. (2003). Cellwise residual analysis in two-way contingency tables. *Educational and Psychological Measurement*, *63*(5), 825-839.

Hinrichs, U., & Carpendale, S. (2011, May). Gestures in the wild: studying multi-touch gesture sequences on interactive tabletop exhibits. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3023-3032). ACM.

Hoggan, E., Nacenta, M., Kristensson, P. O., Williamson, J., Oulasvirta, A., & Lehtiö, A. (2013, October). Multi-touch pinch gestures: Performance and ergonomics. In *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces* (pp. 219-222). ACM.

Jeong, H., & Liu, Y. (2016, September). Computational Modeling of Finger Swipe Gestures on Touchscreen: Application of Fitts' Law in 3D Space. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 60, No. 1, pp. 1721-1725). Sage CA: Los Angeles, CA: SAGE Publications.

Kietrys, D. M., Gerg, M. J., Dropkin, J., & Gold, J. E. (2015). Mobile input device type, texting style and screen size influence upper extremity and trapezius muscle activity, and cervical posture while texting. *Applied Ergonomics*, *50*, 98-104.

Kim, H., Kwon, S., Heo, J., Lee, H., & Chung, M. K. (2014). The effect of touch-key size on the usability of In-Vehicle Information Systems and driving safety during simulated driving. *Applied Ergonomics*, *45*(3), 379-388.

Kim, H., & Song, H. (2014). Evaluation of the safety and usability of touch gestures in operating in-vehicle information systems with visual occlusion. *Applied Ergonomics, 45*(3), 789-798.

Scott, B., & Conzola, V. (1997, October). Designing touch screen numeric keypads: Effects of finger size, key size, and key spacing. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 41, No. 1, pp. 360-364). Sage CA: Los Angeles, CA: SAGE Publications.

Sears, A., Revis, D., Swatski, J., Crittenden, R., & Shneiderman, B. (1993). Investigating touchscreen typing: the effect of keyboard size on typing speed. *Behaviour & Information Technology*, *12*(1), 17-22.

Wolf, K., Schleicher, R., & Rohs, M. (2014, September). Ergonomic characteristics of gestures for front-and back-of-tablets interaction with grasping hands. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services* (pp. 453-458). ACM.

Chapter 3.

Queuing Network Modeling of Touchscreen Drag Gestures using Motion Tracking

3.1. Introduction

As touchscreen technologies have rapidly developed over the last few decades, diverse interfaces that require a wide range of touchscreen gestures have become popular (Saffer, 2008; Bhalla & Bhalla, 2010). Although the definition varies with the research domain, touchscreen gestures can be generally categorized into two types, depending on the number of fingers used: single-touch and multi-touch gestures. Typical examples of a single-touch gesture include tap, swipe, and drag, whereas pinch and spread are examples of a multi-touch gesture.

Since the use of touchscreen gestures has become more prevalent in our daily lives, humans are likely to have more chances to interact with systems or environments using the touchscreen gestures. For instance, they perform touchscreen gestures while walking or driving to navigate to the destination. Thus, it is necessary to investigate human behaviors for touchscreen gesture tasks in the comprehensive process, using human brain and body segments, not just an index finger movement itself.

Many experimental studies have investigated finger gesture performance on touchscreens (e.g., Parhi, Karlson, & Bederson, 2006; Sasangohar, MacKenzie, & Scott, 2009; Jorritsma, Prins, & van Ooijen, 2015; Kim & Jo, 2015; Asakawa, Dennerlein, & Jindrich, 2017, Jeong & Liu, 2017a). In addition to conducting experiments, several studies have developed prediction models using the empirical data obtained through their experiments (e.g., Epps, 1986; Bi, Li, &

Zhai, 2013; Ljubic, Glavinic, & Kukec, 2015). However, most of these models have several limitations. First, they relied on Fitts' law or its extension, treating movement time as a function of only the distance to and size of the target (Fitts, 1954). In other words, these previous models do not consider individual differences that affect finger gesture performance, such as human's anthropometry. Furthermore, the previous models focused on just finger movements, not considering human perceptual and cognitive processes; thus, these models cannot fully represent the details of comprehensive activities in human-computer interaction, such as visual attention shift, and memory storage and retrieval.

According to the literature, the movement time of object (e.g., fingertips or mouse cursors) differs depending on which direction the object is heading to. Jagacinski and Monk (1985) compared the movement times when both joystick and head-mounted sight were used in two-dimensional directions. In their study, horizontal and vertical movement times were slightly shorter than diagonal movement time, but finger movements on touchscreens were not investigated. In addition to movement direction, anthropometry is also regarded as a factor that affects movement time. Bergstrom-Lehtovirta and Oulasvirta (2014) found that hand size was one of the factors to predict the functional area of the thumb on a touchscreen surface. Scott and Conzola (1997) examined the effect of finger size on touchscreen keying performance. They found that finger size had a significant effect on keying speed and duplication errors (i.e., the same digit was inserted into the code again). That is, smaller fingers produced significantly faster keying performance and more duplication errors.

In this chapter, I present a QN-based computational model that predicts finger-drag gesture performance on touchscreen devices, by addressing (1) how to develop "*Drag-with-finger*" operator (because this operator did not exist in the previous QN-models) and (2) how this

new operator plays together with other existing operators in the QN-MHP architecture.

Figure 3.1 shows the conceptual structure for how I modeled single-touch drag gestures in the comprehensive process (i.e., the entire process to complete the finger-drag task, including visual attention shift, memory storage and retrieval, and hand-finger movements). I first needed predictive models in eight different angular directions (i.e., 0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°) to predict how long the movement of the finger-drag performed on the touchscreen devices takes (i.e., the finger movement time of drag-gesture; note that this time corresponds to the "*Drag-with-finger*" operator's execution time). To develop the predictive models, I collected the motion data of finger-drag movement in eight different angular directions, using a motion tracking system. Then I used the time frames of the motion data as a measurement of the finger-drag movement time. Using stepwise regression analysis, I acquired the predictive regression models. The predictors of these regression models in each direction were nine anthropometric parameters: stature (S), finger spread (FS), thumb breadth (HB), index finger breadth (IB), short thumb length (STL), long thumb length (LTL), index finger length (IL), hand length (HL), and hand breadth (HB). Most of parameters were the hand and finger sizes related to the single-touch drag gesture. The definition of each hand-finger anthropometric parameter was adopted from the Eastman Kodak Company (1986) and is illustrated in Figure 3.2. Stature (measured by a linear distance from the floor to the top of the skull) was also selected as an anthropometric parameter because I was to model finger-drag gestures in the assumption of standing position. After the regression equations were developed, using 11 participants' motion tracking and anthropometric data, I used those equations to develop a new operator, "*Drag-with-finger*", and built a comprehensive QN-based model, using other existing operators. As highlighted in Figure 3.1, this new motor operator is associated with not only most of the motor servers (i.e., W-Z and hand

servers) in the QN-MHP, but also a cognitive server, Server D (a procedural long-term memory server that stores the whole task procedure/components, thus all the operators are associated with this server. The details are described in Section 3.2).

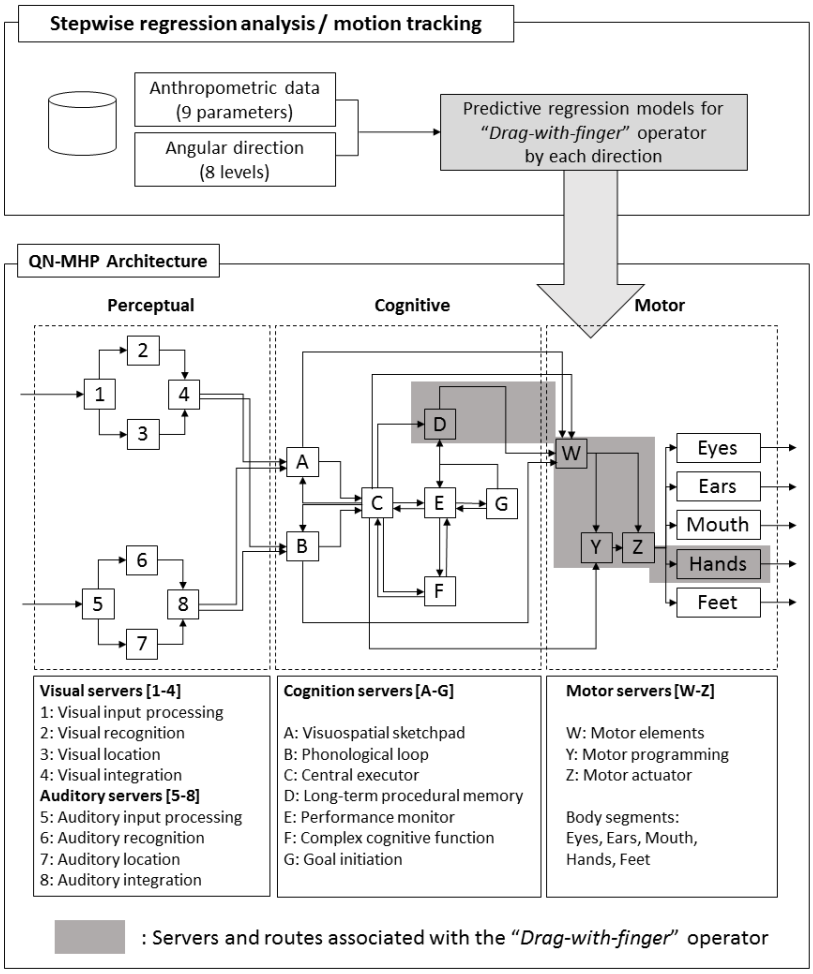


Figure 3.1 – Conceptual structure for the integration of motion tracking and the QN-MHP architecture

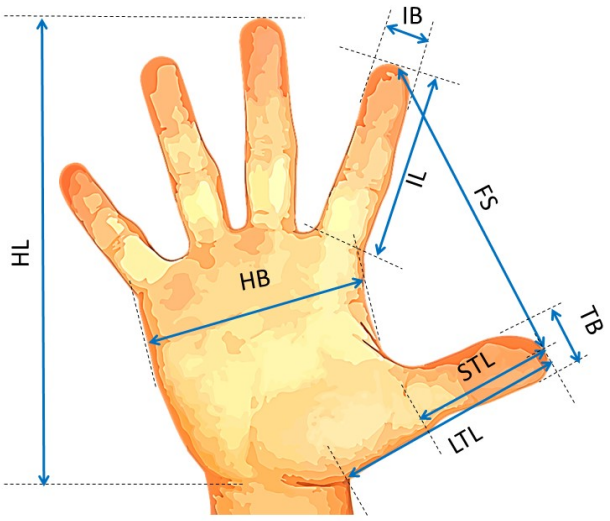


Figure 3.2 – Hand-finger anthropometric dimensions

3.2. Queueing-network modeling of human performance

3.2.1. Task analysis in the QN-MHP Architecture

In the NGOMSL (Natural Goals, Operators, Methods, and Selection rules Language; Kieras, 1999)-style task analysis, task components (TCs) are used to describe each step to accomplish the whole task. Each TC is made with a pre-determined operator that runs with one or multiple parameter(s). Table 3.1 shows the result of task analysis for finger-drag gesture task. In Table 3.1, “Look-at”, “Determine-hand-movement”, “Drag-with-finger” are the examples of the operator, whereas “<target type>”, “<device id>”, “<x, y>” are the examples of the parameter. The comprehensive finger-drag gesture task consisted of 16 task components, describing as three main subtasks.

Table 3.1 – NGOMSL-style task description of comprehensive finger-drag gesture task

Goal: Make a finger-drag gesture on a touchscreen device
TC 1: *Look-at* <target type> on <device id> at location <x, y> TC 2: *Store* <target value> to short-term-memory TC 3: *Retrieve* <target value> from short-term-memory TC 4: *Compare* <target value> to <expected value> If match, go to TC 5, else return to TC 1. TC 5: *Decide*
TC 6: *Look-at* <target type> on <device id> at location <x, y> TC 7: *Store* <target value> to short-term-memory TC 8: *Retrieve* <target value> from short-term-memory TC 9: *Determine-hand-movement* TC 10: *Reach-with-hand*
TC 11: *Look-at* <target type> on <device id> at location <x, y> TC 12: *Store* <target value> to short-term-memory TC 13: *Retrieve* <target value> from short-term-memory TC 14: *Determine-finger-movement* TC 15: ***Drag-with-finger*** with <anthropometric data> in <direction> TC 16: Return with goal accomplished

3.2.1.1. Visual information perception [TCs 1-5]

The first subtask is a process to perceive a visual information and to compare the information to the information expected. It includes to perceive a visual signal (or information) on a touchscreen device with its two-dimension location, and to store and retrieve the information of the visual signal, using short-term-memory. Then, a comparison process is performed to determine whether the information perceived equals to the information expected.

3.2.1.2. Hand reach to the target [TCs 6-10]

In the second subtask, a perception-cognition-motor process using a hand is performed. Specifically, it includes to perceive and remember the information of visual signal's location to reach. Using the information, it is determined whether to move the hand. In the motor subnetwork, the hand servers make the hand be reached to the target, based on the estimation of how far/long the hand reaches the target.

3.2.1.3. Finger drag on the target [TCs 11-15]

In this subtask, similar to the second subtask, a perception-cognition-motor process is performed, but using a finger, not a hand. After receiving and remembering the visual

information to drag, the model determines whether to perform the finger-drag using a finger. The hand servers make the finger to be dragged on the target, based on the estimation of how far/long the finger drags on the target.

3.2.2. Development of operators in the QN-MHP architecture

3.2.2.1. Visual perception operator

Look-at: The purpose of this operator is to allow a human model to look at a specific location. Three parameters are used to set the specific target location: type of target (e.g., text or color), device id, and a target's two-dimensional coordinates on the device. Once the "Look-at" operator is activated at Server D (a long-term procedural memory server), it verifies whether an eye-saccade motor action is needed, which is used to make the current visual attention be on the target location. If the saccade action is needed, because the current visual attention is not on the target location, it triggers a saccade motor action at Server W (a motor-elements server). Then Server W triggers the Eyes server so that a saccade can be executed at the Eyes server. The saccade execution time is determined by a visual angle (i.e., the angle from the current location of visual attention to the target location) and an angular velocity (i.e., 4 msec/degree; Kieras & Meyer, 1997). After the saccade is completed at the Eyes server, an entity (or visual stimulus) of target enters into Server 1 (a visual input server). Then the entity enters Servers 2 (Visual recognition) and 3 (Visual location) in parallel, and thus the human model can recognize the target and its location. Through Server 4 (a visual integration server), the entity is transformed into cognitive subnetwork including Servers A-G. For the finger-drag task in this study, the geometric center of a virtual touchscreen device was set as the location of the target to look at. On the target, a green circle was used as a visual stimulus in this model. Since the initial visual attention was set on the device, so the saccade motor action was not required.

3.2.2.2. Memory and cognition operators

Store to short-term-memory: Using this operator, the model stores the target information to the short-term-memory through Server A or B, depending on the type of the target information. Server A stores visual and spatial information, whereas Server B does phonological information. Once this operator is activated at Server D, the entity enters Server A or B. For the finger-drag task in this study, the entity went to Server B (a phonological-loop server), because a text was used as a visual stimulus, not color or other types.

Retrieve from short-term-memory: This operator allows the model to retrieve target information from the short-term-memory and makes the retrieved value be available for other operators. Similar to the "Store to short-term-memory" operator, once this operator is activated at Server D, the entity enters Server A or B. Then, the entity would enter Server C (a central-executor server), if operators for cognitive central executions (e.g., "Compare" or "Compute") follows after this "Retrieve from short-term-memory". For the finger-drag task in this study, Server B was used to retrieve the text information.

Compare: The model compares the target and expected values, using this operator. The type of information can be either text or color. If the values are identical, the result of 1 is returned (going to the next TC), otherwise, 0 is returned (going back to the TC of visual perception). Once this operator is activated at Server D, the entity enters Server C to (1) either only conduct the comparison if the information type is color; (2) or further route the incoming cognitive entity to Server F (a cognitive-complex function server) if the information type is text. In this finger-drag task, Server C was used, because a green circle was used as a visual stimulus.

Determine-hand-movement / *Determine-finger-movement*: These operators allow the model to determine whether to move the hand and finger. Once these operators are activated at

Server D, the entity enters Server C and the model determines whether or not to move their hand and finger. Then, Server C would route the entity to Server W (a motor-elements server), if operators for motor executions (e.g., "Reach-with-hand" or "Click-with-finger") follow after these "Determine-hand-movement" or "Determine-finger-movement" operators.

3.2.2.3. Motor operators

Reach-with-hand: This operator initiates a reaching action using the model's hand servers. Once this operator is activated at Server D, a motor entity is created in Server W with the motor type of "Reach-with-hand". This motor entity is then processed in Servers W, Y, Z, and Right-hand or Left-hand servers. The hand servers make the hand be reached to the target, based on the estimation of how far/long the hand reaches the target. The general Fitts' law equation is used to determine the reaching execution time. According to Shannon formulation (MacKenzie & Buxton, 1992), the movement time MT is:

$$MT = a + b \times \log_2(\frac{A}{W} + 1) \qquad (1)$$

, where *a* and *b* are empirical regression coefficients, varying in the environment (e.g., people and devices). W refers to the target's effective size, whereas A refers to the distance to the target. According to Wobbrock, Cutrell, Harada, and MacKenzie (2008), the effective size of W (W') is:

$$W' = \min(Height, Width) \times \frac{Z_{Fitts}}{W} \qquad (2)$$

and 4% of error rate is assumed in Fitts' law. Since the MacKenzie and Buxton (1992)'s original model included a mouse clicking time, which corresponds to 260 msec, according to ACT-R 6.0 model, this amount of time was subtracted. The 70 msec of motor preparation time was also deducted to determine the execution time of the "*Reach-with-hand*" operator. For the finger-drag task in this study, it is assumed that the model uses a right hand to reach the target. Also, it is

assumed the reaching distance is 300 mm, which closely resembles the actual distance found from the pilot test. a = 230 (msec), b = 166 (msec/bit) was used from MacKenzie and Buxton (1992). Each participant's index finger breadth (i.e., IB × IB) was set as the target's height and width in this study.

Drag-with-finger: This operator initiates a dragging action using the model's hand servers. Once this operator is activated at Server D, a motor entity is created in Server W with the motor type of "*Drag-with-finger*". Similar to the "*Reach-with-hand*" operator, this motor entity is then processed in Servers W, Y, Z, and Right-hand or Left-hand servers. The hand servers make the finger be dragged on the target, based on the estimation of how far/long the finger drags on the target. Note that this motor operator is directly associated with only TC 15, not any other TCs. In this study, the dragging execution time was determined by the regression models derived from hand-finger-related anthropometric data and eight different angular directions (i.e., the outcomes from Section 3.3.5). The regression equations for the finger movement time of drag-gesture are shown in Table 3.2.

Table 3.2 – Prediction equations for the finger movement time of drag-gesture

Angular direction (degree)	Prediction equations (milliseconds)	Adjusted R^2	RMSE	M ± SD (milliseconds)
0	1431 - 2610 *TB* + 3111 *IB*	.70	.23	1127 ± 422
45	2552 - 2643 *TB* + 649 *STL*	.59	.40	1246 ± 621
90	62 - 2358 *TB* + 3662 *IB*	.52	.34	1177 ± 490
135	3375 -3191 *TB* + 697 *STL*	.55	.50	1252 ± 759
180	1208 - 2153 *TB* + 2615 *IB*	.68	.20	1030 ± 361
225	2963 - 2547 *TB* + 531 *STL*	.57	.39	1114 ± 590
270	1171 - 2618 *TB* + 2349 *IB* + 107 *FS*	.71	.23	1147 ± 435
315	- 1749 - 4508 *TB* + 4289 *IB* + 768 *IL*	.56	.64	1573 ± 973

3.2.2.4. Procedural flow operators

Decide: It specifies the procedural sequences of the steps in a list of task. Once this operator is activated at Server D, it makes the task steps specified in its parameters depending on the result of the previous step.

Goal-accomplished: This operator indicates the completion of a task. Once this operator is activated at Server D, it verifies whether there is any other pending task, and switches to that task if any task remains. If no other pending task is remaining, this operator allows the model to terminate the simulation.

3.3. Motion-tracking data collection for "Drag-with-finger" operator development

3.3.1. Participants

Eleven participants (6 males and 5 females) ranging in age between 20 and 30 years (M = 25.5, SD = 2.9) were recruited from the campus at the University of Michigan. All participants had normal or corrected-to-normal vision and were right-handed. They reported no physical issues in using touchscreen display and used touchscreen devices (e.g., smartphones and tablets) for 7.5 years. Participants were paid for their time with $15 hourly rate in cash. Table 3.3 summarizes anthropometric data obtained from the participants, and population percentile data extracted from Greiner (1991) to compare with the participants' data.

Table 3.3 – Summary of anthropometric data for the "*Drag-with-finger*" operator development (n = 11)

No.	Anthropometric Dimension (unit: millimeters)	M	SD	Min	Max	Population Percentiles (male/female)		
						5th	50th	95th
1	Stature (S)	1743	88	1622	1857	1651/1527	1755/1629	1868/1738
2	Finger spread (FS)	146	22	111	171	unknown	unknown	unknown
3	Thumb breadth (TB)	21	2	17	24	22/19	24/21	26/23
4	Index finger breadth (IB)	16	2	14	19	18/15	20/17	23/19
5	Short thumb length (STL)	64	8	55	74	62/56	70/63	78/72
6	Long thumb length (LTL)	128	10	109	145	124/112	138/125	153/141
7	Index finger length (IL)	73	6	63	84	67/62	75/70	84/77
8	Hand length (HL)	187	14	166	210	179/163	194/178	212/195
9	Hand breadth (HB)	84	7	73	95	86/76	95/83	105/90

3.3.2. Apparatus

A motion tracking system (OptoTrak® Certus™; Northern Digital Inc.) with two standing position sensors (3 cameras on each sensor; 3.5 m away from each other) was used to record finger movements for finger-drag gestures. One marker was attached on the center of participants' right index fingernail and it was secured with Velcro® straps across the finger, wrist, and forearm, shown in Figure 3.3. A touchscreen device (iPad; 1024 x 768; 132 ppi; 9.7-inch LED-backlit glossy widescreen Multi-Touch display) was mounted on the table (Height = 95 cm). Participants were asked to find the most comfortable standing position so they do not feel any discomfort while performing the finger-drag gesture tasks.

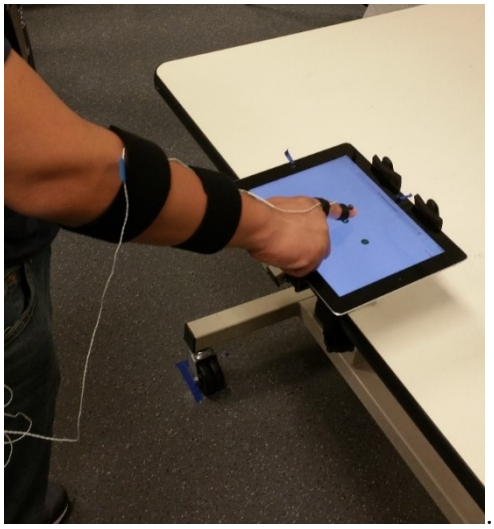


Figure 3.3 – Experimental setup of the finger-drag gesture task

3.3.3. Touchscreen gesture task and experimental design

The performance of finger-drag gestures was measured, using a touchscreen interface used in Jeong and Liu (2017a). As shown in Figure 3.4, nine circles (i.e., eight target circles around one center circle) were designed on a touchscreen display, but only two circles (i.e., one center circle and one of the eight target circles; colored in green – one with a hole, the other

without a hole) were presented to the participants during the experiment. The distance between the center and the target circles was 40 mm, a fixed value. Participants were instructed to move their right index fingers from the center circle to one of the target circles; while the center circle was fixed and always presented, the target circle was randomly presented on the touchscreen display. The radius of the both center and target circles was identically 20 pixels (= 9 mm). Whenever the finger arrives from the air to the display surface and leaves from the display surface to the air, a 20-pixel-radius black circle was shown on the display, as a visual feedback. Only when the Euclidean distance between the center of the black and the green circles equals to or less than 20 pixels (also called a match of the black and green circles), it was defined as a success. The participants were asked to press 'start' button on the center of the screen (then the button disappeared immediately), and then to complete the dragging task with two successes on the center and target circle matches (i.e., initial and final matches). In the current study, only the data of the success trials were used and analyzed. In other words, I did not model the accuracy of the drag-gesture's performance. Instead, I used and analyzed time data, only for the success trials.

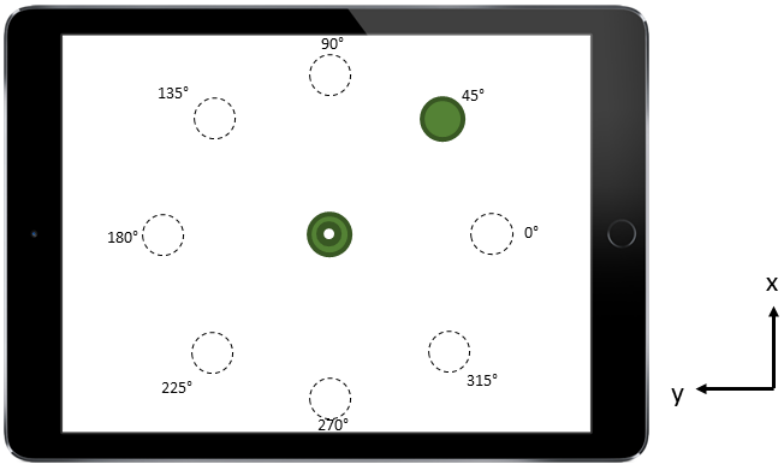


Figure 3.4 – Angular directions for finger-drag gesture and the coordinate system
(Note: This figure shows an example task when the direction is 45°; the dotted circles and the degree information were not presented to the participants)

A within-subject factorial design was used in this study. Two independent variables (including a subject variable) manipulated in this experiment were 8 different angular directions (i.e., 0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°) and the participants' 9 anthropometric parameters (i.e., S, FS, HB, IB, STL, LTL, IL, HL, and HB). Each of eleven participants conducted 3 replications of the drag gesture to each angular direction. Each participant performed finger-drag gestures in two sessions: 1) horizontal (0° and 180°) / vertical (90° and 270°) and 2) diagonal direction (45°, 135°, 225°, and 315°). The order of conditions was balanced across participants.

3.3.4. Experimental procedures

After arrival in the laboratory, each subject was informed of the purpose and nature of the experimental task. They completed a consent form and filled out a brief survey about their demographic information (e.g., age and gender) and years experienced of touchscreen usage. After each participant's anthropometric data (e.g., hand and finger size) were measured, a marker and motion-tracking system wires were attached to their right index finger and upper extremity. Prior to the actual finger-drag gesture task, they practiced sample trials to get familiar with the tasks for 10-15 minutes. Each of eleven subjects conducted two sessions (i.e., horizontal/vertical and diagonal) for drag gesture tasks with their right index finger. At the end of the experiment, they completed a payment form and were paid for their times, taken for approximately less than an hour.

3.3.5. Development of finger-drag movement time regression models for "Drag-with-finger" operator

Out of 264 motion data sets (i.e., 11 participants × 8 directions × 3 replications), there were only 6 missing data points caused by procedural and equipment malfunction (including non-

success trials). Figure 3.5 shows a subject's sample motion data, depending on the angular directions. The time frames (in milliseconds) for each motion (in millimeters) were collected and they were regarded as the execution time of "*Drag-with-finger*" operator. SPSS Statistics 23 was used for stepwise regression analysis to model the relationship between the finger-drag movement time and potential predictors including all anthropometric parameters. The criteria were set as probability-of-F-to-enter ≤ .05 and probability-of-F-to-remove ≥ .10. In order to detect the multicollinearity problem (i.e., having highly correlated predictors in a regression model), variance inflation factors (VIFs) was used to check whether it is less than 10 (Kutner, Nachtsheim, & Neter, 2004; Montgomery, Peck, & Vining, 2015). High degree of multicollinearity was not present for all eight conditions. From the automated stepwise process, linear regression equations were obtained depending on each angular direction (see Table 3.2).

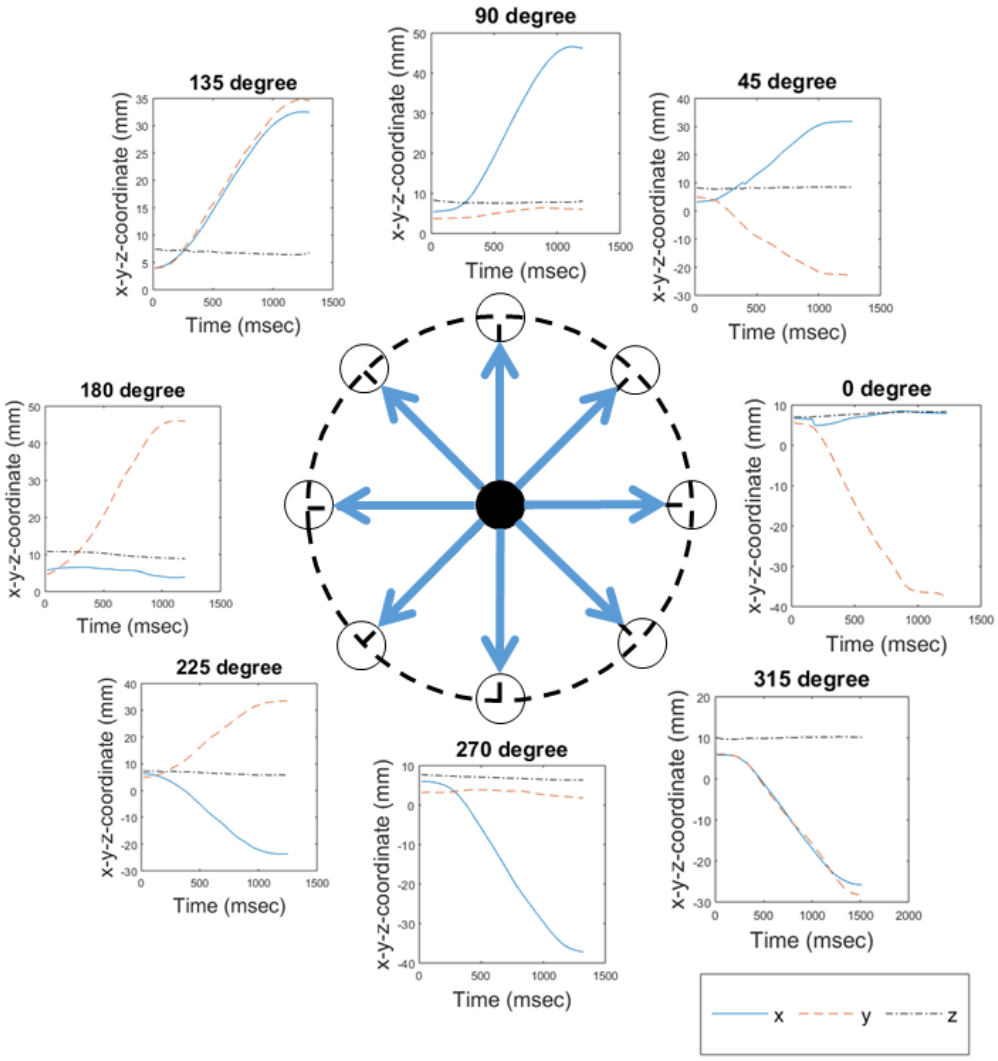


Figure 3.5 – A sample motion data in eight different angular directions

3.4. Model validation

To validate the model including (1) finger-drag gesture process (i.e., TC 15 only) and (2) its comprehensive process (i.e., all TCs 1-16), data from Chapter 2 (Jeong & Liu, 2017a) were used. The data were collected from twenty (11 males and 9 females; age M = 23.8, SD = 3.3) right-handed students who had used the touchscreen devices for 6.7 years on average. Participants' anthropometric data are shown in Table 3.4 (Note that the data were collected in Jeong & Liu (2017a), but not used in that study). The touchscreen software collected the finger

movement time of drag-gesture (i.e., only the motion time of the finger touched and dragged on the surface of touchscreen) and the comprehensive process time (i.e., the entire process time to complete the finger-drag task, from when 'start' button was pressed to when the task was completed).

Using the new operator "*Drag-with-finger*" with other existing operators (e.g., "*Look-at*", "*Reach-with-hand*", and "*Compare*"; adopted from the previous studies (Feng, Liu, Chen, Filev, & To, 2014; Jeong & Liu, 2017b), the QN-based model was run using the comprehensive process shown in Table 3.1. The model performed 20 simulation runs with different angular directions in the MATLAB-Simulink software. Figure 3.6 shows the modeling results (i.e., solid lines) compared with experimental results (i.e., dashed lines) in both finger movement time and a comprehensive process time of drag-gesture. In the comparison of finger-drag gesture process time, the R^2 of the model was .90 and the RMS = 92.2 msec; in the comprehensive process time, the R^2 of the model was .80 and the RMS = 256.4 msec.

Table 3.4 – Summary of anthropometric data for validation (n = 20)

No.	Anthropometric Dimension (unit: millimeters)	M	SD	Min	Max
1	Stature (S)	1729	98	1562	1869
2	Finger spread (FS)	149	21	111	185
3	Thumb breadth (TB)	20	2	17	24
4	Index finger breadth (IB)	16	1	14	19
5	Short thumb length (STL)	64	7	55	79
6	Long thumb length (LTL)	127	10	109	145
7	Index finger length (IL)	73	6	63	86
8	Hand length (HL)	186	13	165	210
9	Hand breadth (HB)	84	7	73	99

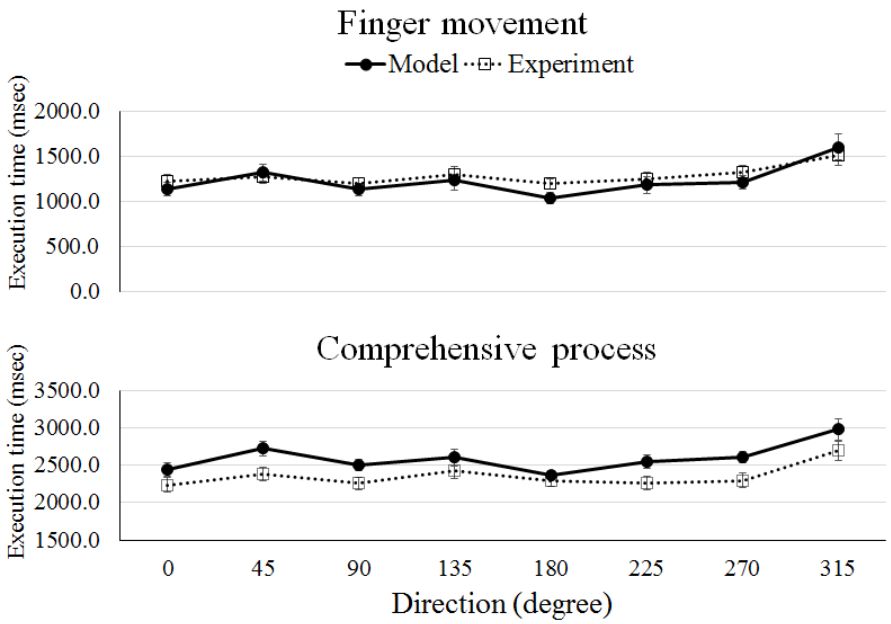


Figure 3.6 – Execution times in drag-gesture's finger movement (up) and comprehensive process (bottom) time

3.5. Discussion

This chapter presented a computational model for finger-drag gestures on touchscreen devices, integrating the QN-MHP architecture and motion tracking: specifically, for finger movement and comprehensive process of the drag-gesture. In addition to developing the QN-based model for the touchscreen finger-drag gesture, the model was evaluated with the data from twenty participants in Chapter 2 (Jeong & Liu, 2017a). The model was able to generate similar outputs (R^2 was more than 80% and RMS was less than 300 msec.) of the gesture execution times with human performance data.

3.5.1. Implications

Since the QN-MHP framework was first developed, a wide range of human performance has been successfully modeled (e.g., map reading (Liu et al., 2006), transcription and numerical typing (Wu & Liu, 2008; Lin & Wu, 2012), and driving control (Bi, Gan, Shang, & Liu, 2012; Jeong, Feng, & Liu, 2017)). The application domains have been expanded to more practical

fields, but there has been no attempt to model the touchscreen finger-drag gesture performance. In addition to the fact that this study is the first attempt to model the finger-drag gesture performance, it has two significant features, compared to other previous modeling studies of touchscreen gestures (e.g., Epps, 1986; Bi, Li, & Zhai, 2013; Ljubic, Glavinic, & Kukec, 2015). First, this study considered the user's hand anthropometric data as a factor that affects the finger-drag performance. From the stepwise regression analysis, it turned out that the finger-drag movement time depends on its direction, and it is a function of index finger breadth (IB), thumb breadth (TB), short thumb length (STL), index finger length (IL), and finger spread (FS), shown in Table 3.2. Using the predictive regression models with the angular direction and several major anthropometric data, the movement time of finger-drag can be predictable. The anthropometry-based prediction model of movement time is a novel approach in finger-drag study, in contrast with the conventional way, using the target size and distance. Using the anthropometric data under the queueing network cognitive framework, this study predicted not only the simple finger-drag gesture movement time but also the comprehensive process time. Since most tasks, in reality, are more likely to be performed in multitask settings, it is necessary to understand and predict both the simple task process (e.g., a single finger-drag across a touchscreen) and more comprehensive process (e.g., additional behaviors of looking at a target on the touchscreen and determining the hand and finger movements).

Moreover, the model developed in this chapter can provide potential benefits to usability test researchers who investigate the drag-gestures on certain touchscreen products. For example, the model can help minimize the cost of usability evaluation for touchscreen interfaces using finger-drag gestures. At the early stage of product design (i.e., prior to developing the prototypes or products), the model can be used to compare the human performance such as task completion

time, depending on the gesture direction and user's hand-finger anthropometric data. Furthermore, this finger-drag model can be integrated with other touchscreen gesture models (e.g., Jeong & Liu, 2016, 2017c) in order to predict human performance while conducting tasks using multiple touchscreen gestures, such as swiping, pinching, and spreading.

3.5.2. Limitations and future studies

Although this research successfully integrated a queueing network model with studies on the motion tracking technology, it shows some limitations. First, the current model only considers the anthropometry and drag-gesture direction as predictors of the finger-drag gesture performance in the condition of fixed target size and distance. Although I determined these two predictors from literature review as mentioned in the introduction, there might be some other factors that may need to be taken into account in future studies, such as the handheld postures (i.e., one hand or two hands), multitasking conditions (i.e., while walking or driving), as well as various target sizes and distances. Second, while the anthropometric data related to the "*Drag-with-finger*" operator were used, the anthropometric data for the "*Reach-with-hand*" operator were not used in the current study. Thus, I will further expand the current model to consider not only the anthropometric data of hand and finger sizes, but also other data including upper and lower arm lengths. Third, the parameters used for the "*Reach-with-hand*" operator (i.e., a = 230 ms, b = 166 ms/bit) were originally developed for the mouse-based point-and-click task in MacKenzie and Buxton (1992), that might not be suitable for the touchscreen-based touch-and-drag task used in the current study. Therefore, further study is needed to investigate more appropriate empirical data for the "*Reach-with-hand*" operator for better model validation performance. Fourth, the current model in this study has only focused on predicting performance times, without considering other performance outcomes, such as error rate in touchscreen finger-

input tasks. Further work can enrich the scope of the model so that the model can predict a wider range of human performance in complex tasks. Lastly, the anthropometric data used in this study might be categorized as a nested data (i.e., fingers nested in hands), referring to the data that are organized by a hierarchical or multilevel structure with more than one level (Aarts et al., 2014). Although the multicollinearity test was performed prior to applying the stepwise regression analysis in this study, the dependency among finger and hand dimensions could exist. Thus, further study can apply a multilevel analysis with more data points for each parameter to investigate whether the outcomes addressed in this study were over-fitted, by comparing the accuracy rate between the two analyses (Cohen et al., 2003; Burnham & Anderson, 2003; Harrell, 2015).

References

Aarts, E., Verhage, M., Veenvliet, J. V., Dolan, C. V., & Van Der Sluis, S. (2014). A solution to dependency: using multilevel analysis to accommodate nested data. *Nature Neuroscience*, *17*(4), 491-496.

Asakawa, D. S., Dennerlein, J. T., & Jindrich, D. L. (2017). Index finger and thumb kinematics and performance measurements for common touchscreen gestures. *Applied Ergonomics*, 58, 176-181.

Bergstrom-Lehtovirta, J., & Oulasvirta, A. (2014). Modeling the functional area of the thumb on mobile touchscreen surfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1991-2000). ACM.

Bhalla, M. R., & Bhalla, A. V. (2010). Comparative study of various touchscreen technologies. *International Journal of Computer Applications*,*6*(8), 12-18.

Bi, L., Gan, G., Shang, J., & Liu, Y. (2012). Queuing network modeling of driver lateral control with or without a cognitive distraction task. *IEEE Transactions on Intelligent Transportation Systems*, *13*(4), 1810-1820.

Bi, X., Li, Y., & Zhai, S. (2013). FFitts law: modeling finger touch with fitts' law. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1363-1372). ACM.

Burnham, K. P., & Anderson, D. R. (2003). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.

Eastman Kodak Company, Ergonomics Group (1986). Ergonomic design for people at work. Vol. 1. *Van Nostrand Reinhold*, New York, 28-34.

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

Feng, F., Liu, Y., Chen, Y., Filev, D., & To, C. (2014). Computer-Aided Usability Evaluation of In-Vehicle Infotainment Systems. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, No. 1, pp. 2285-2289). SAGE Publications.

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

Greiner, T. M. (1991). *Hand Anthropometry of US Army Personnel* (No. TR-92/011). Army Natick Research Development and Engineering Center MA.

Harrell, F. (2015). *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer.

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

Jeong, H., Feng, F., & Liu, Y. (2017). Computational Modeling of Driver Lateral Control on Curved Roads with Integration of Vehicle Dynamics and Reference Trajectory Tracking. In *Proceedings of the Ninth International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (pp.193-199), Manchester Village, VT.

Jeong, H., & Liu, Y. (2016). Computational Modeling of Swipe Gestures on Touch Screens: Application of Fitts' Law in 3D Space. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 60, No. 1, pp. 1721-1725). SAGE Publications.

Jeong, H., & Liu, Y. (2017a). Effects of touchscreen gesture's type and direction on finger-touch input performance and subjective ratings. *Ergonomics*. 60(11), 1528-1539.

Jeong, H., & Liu, Y. (2017b). Modeling of Stimulus-Response Secondary Tasks with Different Modalities While Driving in a Computational Cognitive Architecture. In *Proceedings of the Ninth International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (pp.58-64), Manchester Village, VT.

Jeong, H., & Liu, Y. (2017c). Modeling Touchscreen Pinch and Spread Gestures in a Computational Cognitive Architecture. In *Proceedings of the 67th Annual Conference and Expo of the Institute of Industrial and Systems Engineers (IISE)*, Pittsburgh, PA, 2171-2176.

Jorritsma, W., Prins, J. T., & van Ooijen, P. M. (2015). Comparing four touch-based interaction techniques for an image-based audience response system. *International Journal of Human-Computer Interaction*, *31*(6), 440-450.

Kim, I., & Jo, J. H. (2015). Performance comparisons between thumb-based and finger-based input on a small touch-screen under realistic variability. *International Journal of Human-Computer Interaction*, *31*(11), 746-760.

Kieras, D. E., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-computer interaction*, 12(4), 391-438.

Kieras, D. E. (1999). A guide to GOMS model usability evaluation using GOMSL and GLEAN3. *University of Michigan*, (313).

Kutner, M. H., Nachtsheim, C., & Neter, J. (2004). *Applied linear regression models*. McGraw-Hill/Irwin.

Lin, C. J., & Wu, C. (2012). Mathematically modelling the effects of pacing, finger strategies and urgency on numerical typing performance with queuing network model human processor. *Ergonomics*, *55*(10), 1180-1204.

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.

Ljubic, S., Glavinic, V., & Kukec, M. (2015). Finger-Based Pointing Performance on Mobile Touchscreen Devices: Fitts' Law Fits. *In International Conference on Universal Access in Human-Computer Interaction* (pp. 318-329). Springer International Publishing.

MacKenzie, I. S., & Buxton, W. (1992). Extending Fitts' law to two-dimensional tasks. *Proceedings of the CHI 1992 Conference on Human Factors in Computing Systems, ACM, New York*, 219-226.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2015). *Introduction to linear regression analysis*. John Wiley and Sons.

Parhi, P., Karlson, A. K., & Bederson, B. B. (2006). Target size study for one-handed thumb use on small touchscreen devices. In *Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 203-210). ACM.

Saffer, D. (2008). *Designing gestural interfaces: Touchscreens and interactive devices*. O'Reilly Media, Inc.

Sasangohar, F., MacKenzie, I. S., & Scott, S. D. (2009). Evaluation of mouse and touch input for a tabletop display using Fitts' reciprocal tapping task. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 53, No. 12, pp. 839-843). SAGE Publications.

Scott, B., & Conzola, V. (1997). Designing touch screen numeric keypads: Effects of finger size, key size, and key spacing. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 41*(1), 360-364.

Wobbrock, J. O., Cutrell, E., Harada, S., & MacKenzie, I. S. (2008). An error model for pointing based on Fitts' law. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1613-1622). ACM.

Wu, C., & Liu, Y. (2008). Queuing network modeling of transcription typing. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 15(1), 6.

Chapter 4.

An Experimental Study on the Effect of In-vehicle Secondary Task Modalities and Road Curvature on Eye Movements and Driving Performance

4.1. Introduction

Crash statistics indicate the crash rates on horizontal curves, referring to laterally bent roads designed to connect to straight roads, are three times higher than on straight road segments (Glennon, Neuman, & Leisch, 1985). Moreover, a report that analyzed crash data from the Fatality Accident Reporting System for the period 1991–2007, found that the odds of being involved in a fatal crash on a curved road are 1.74 times greater than on straight roads (Liu & Subramanian, 2009). In this regard, car crashes on horizontal, curved roads have been considered a significant safety concern for several decades (Johnston, 1982; Zegeer et al., 1991; Charlton, 2007). Many studies have defined risk factors for horizontal curve accidents, such as road alignments, distraction, speeding, alcohol, fatigue, and adverse weather conditions. According to the literature, the most critical factors include: (1) driver distraction by secondary tasks (Vollrath & Trotzke, 2000; Dingus et al., 2006; Young, Lee, & Regan, 2008; Benedetto, Calvi, & D'Amico, 2012) and (2) unsafe road conditions, including sharp road curvatures (Milton & Mannering, 1998; Caliendo, Guida, & Parisi, 2007; Fitzpatrick, Lord, & Park, 2010; Calvi, 2015).

To date, several studies have investigated curve-driving performance dependent on either the different modalities of distraction tasks or on different road curvatures. However, few studies have investigated the combined effects of both horizontal road curvature and secondary-task

sensory modality. For example, Hibberd, Jamson, and Carsten (2013) used six different modality types, three stimuli (visual, auditory, haptic) × two responses (manual and speech), to examine the effects of distraction on driving performance. They concluded that brake reaction time increased when drivers were distracted, but a specific comparison between the different modalities of the distractions was not investigated. Angell et al. (2006) compared the effects of visual-manual and auditory-speech tasks on lane-keeping performance and eye movement. Based on the results of the auditory-speech tasks, there was a longer duration of glances at the road, and the total number of glances allowed drivers to maintain their eyes and attention on the road, compared to visual-manual tasks. However, the effect of different road environments was not examined. Another driving simulation study (Kountouriotis & Merat, 2016) investigated the effects of visual (using a visual search task) and non-visual (using a counting backwards task) distractions on steering performance and eye movement pattern using two different road types: curved and straight. Although they concluded that the effects varied between straight and curved roadways, the effect of road curvature was not examined. Tsimhoni and Green (2001) studied the effects of road curvature on visual demand and driving performance using three curvature radii (194, 291, and 582 m) conditions and visual-speech tasks. They found that the drivers' visual demands increased for sharper curves leading to poorer driving performance. However, the various modality types of the secondary tasks were not specified. Furthermore, Benedetto et al. (2012) investigated the effects of three different ways of using a cellular phone (i.e., hands-held, hands-free, and hands-free voice systems) while driving in three different road scenarios (i.e., a motorway, rural, and urban scenarios). However, the modalities used to operate the cellular phones were not specifically compared.

In summary, few previous studies have investigated diverse secondary-task modalities and multiple road curvatures in combination. Studies of their combined effects can provide not only insights for future design of in-vehicle systems but also implications for design policies and guidelines in order to reduce the frequency of traffic accidents on horizontal, curved roads. For this purpose, the current study was carried out to investigate the impact of both secondary-task modality and road curvature on safe driving for various levels of modality types and road curvatures.

4.2. Literature review

When conducting experiments in dual-task settings (i.e., driving and secondary tasks), how participants are instructed to prioritize their performance between the two tasks is important, because the instruction can result in different outcomes. Janssen and Brumby (2010) conducted a driving simulation experiment using a secondary phone-dialing task while driving. The participants in the study were instructed to prioritize either faster dialing (i.e., dialing-focused) or safer driving (i.e., steering-focused). Under the dialing-focused condition, they glanced longer at the phone and inputted more digits at one time. In contrast, under the steering-focused condition, they made shorter glances away from the road and tried to maintain good driving performance. Horrey, Wickens, and Consalus (2006) conducted a driving experiment using in-vehicle technologies (IVTs), and they also found a significant effect of task priority (i.e., driving-focused, IVT-focused, and equal-focused) on lane-keeping performance and eye movement. When driving was prioritized, variability in lane keeping was lower and the percentages of dwell time on the road were greater, compared to other conditions.

In the current study, participants were instructed to consider driving safety (i.e., maintaining a vehicle as close as possible to the center of the driving lane) as having a higher priority over secondary task performance (i.e., completing stimulus-response tasks as accurately as possible), because prioritizing driving safety is more natural in real life situations. That is, most drivers tend to try to save their lives when there is any conflict between life-threating risk and in-vehicle system operation.

4.2.1. Effects of secondary-task modality

The 100-Car Naturalistic Driving Study (Dingus et al., 2006) found that more than half of all crashes involved secondary-task distractions. The majority of the secondary-task distractions were a result of driver-vehicle interaction, such as using wireless devices (e.g., talking, listening, and dialing) and performing vehicle-related tasks (e.g., adjusting the radio and navigation). Drivers consistently interact with in-vehicle systems to achieve goals, such as maintaining safe driving performance or obtaining particular information for a task (e.g., operating a GPS to navigate to a particular destination). In general, the driver-vehicle interaction represents a type of stimulus-response task: once drivers receive information (or perceive stimuli) from a vehicle system, they respond to the system with a method that is recognized by the system. In past years, in-vehicle systems used only one modality “visual” and one input “manual”, however, more recent systems include additional modalities (e.g., auditory) and inputs (e.g., speech) to maximize the efficiency of driver-vehicle interaction. For example, in the past, while tuning a car radio, drivers read the current frequency (e.g., 97.0 MHz as a visual stimulus) and pressed a button (e.g., Tune + as a manual response) to reach a target frequency (e.g., 97.1 MHz). However, nowadays, to tune a car radio, drivers can listen to a voice announcement (e.g., “Say a frequency” as an auditory stimulus) from the system and make a voice command (e.g., “ninety-seven point one” as a speech response).

Stimulus-response (S-R) compatibility usually refers to how natural the association between the stimulus and the response modality is. In general, response time to a stimulus is faster for a higher level of S-R compatibility. S-R compatibility level is higher when the S-R types use a similar type of modality. For example, an auditory-speech type of task requires a shorter response time than an auditory-manual type of task (Stelzel & Schubert, 2011). As an extension of S-R compatibility study, Stimulus-central-response (S-C-R) compatibility was introduced in the 1980s, emphasizing additional processing that occurs in the brain (Wickens, Sandry, & Vidulich, 1983; Wickens, Vidulich, & Sandry-Garza, 1984). With complex systems, response to a stimulus is likely to be postponed because of the central-brain processing for decision-making. Central processing fundamentally varies, with two different types of information, based on how the working memory operates: spatial or verbal. Spatial information indicates any information associated with geometric objects, whereas, verbal information relates to language, numeric, or symbolic codes. Wickens et al. (1983) found that response-task performance was best for auditory-stimulus speech response (A-S) type and worst for visual-stimulus manual response (V-M) type, and the visual-stimulus speech response (V-S) and auditory-stimulus manual response (A-M) types are in between without large differences when verbal information was used. In contrast, performance with spatial information was reversed across the different modalities. In addition, when dual tasks were conducted, time-sharing efficiency is considered as the extent of how much operators can multitask. In general, the S-R modality with higher performance for a single task had higher time-sharing efficiency for dual tasks.

While performing secondary tasks, drivers tend to have higher visual demand when they look at visual stimuli compared to when they listen to auditory stimuli. On the other hand, drivers tend to have higher visual demand when they respond manually, rather than verbally, because they

need to look at which buttons to press. In addition, it is assumed that once drivers complete the secondary tasks, their visual attention immediately returns to the road to concentrate on driving, if they are instructed to prioritize driving safety over secondary-task performance. Therefore, it is hypothesized that if drivers have higher visual demand because of secondary tasks on curved roads then they will have more unsafe and unstable eye movements for driving (e.g., more fixations off the road) (H1).

Furthermore, many previous studies have investigated the effect of driver inattention on lane-keeping performance (e.g., Liang & Lee, 2010; Peng, Boyle, & Hallmark, 2013). One of the typical findings was that longer eyes-off-road time caused more inconsistent steering control and larger lane deviations. In addition, Klauer, Dingus, Neale, Sudweeks, and Ramsey (2006) found that longer durations of eyes-off-the-forward-roadway led to more near crashes or incidents, because of larger lane deviations. Specifically, total eyes-off-road time (TEORT) greater than two seconds significantly increased crash risk, whereas, an eye-glance duration of less than two seconds did not significantly increase risk compared to normal driving. Moreover, Horrey et al. (2006) developed a regression model for predicting lower variability in lane keeping by higher percentages of dwell time on the road. Thus, it is hypothesized that when drivers have higher visual demand from secondary tasks on curved roads then they will have greater variability in lane-keeping performance (H2).

4.2.2. Effects of road curvature

Several studies have found that the degree of road curvatures is to be one of the most influential geometric factors associated with safety on horizontal, curved roads. McGee (2011) mentioned that majority of states in the United States are concerned with the horizontal curvature as geometric design criteria, and this horizontal curvature affects other primary criteria, including

speed and stopping sight distance. Milton and Mannering (1998) found a strong correlation between accident rate and the radius of the horizontal curve. That is, a decrease in curve radii increased the frequency of accidents. Moreover, it was found that the dispersion of trajectory, which refers to a vehicle's displacement within a driving lane, varied with different road curvatures (Calvi, 2015).

Previous driving studies have found that the degree of a driver's visual attention to secondary tasks varied depending on road conditions. For example, Kun, Brumby, and Medenica (2014) found that drivers took shorter glances at in-vehicle devices for secondary tasks while driving in the city, compared to the highway. In addition, one finding in Liu (2003) was that drivers took shorter times to detect speed limit signs on the side of the road when they were driving in scenarios that had greater driving load (i.e., more curved roads). Similarly, in Charlton's (2007) study, it was found that lane-keeping on curved roads requires more visual attention than on straight roads. McDonald and Ellis (1975) found that straight roads required 23% of a driver's attentional demand, whereas, curved roads required higher levels of a driver's attentional demands (26%). Under the assumption that driving safety is prioritized over secondary-task performance in harder driving conditions, it is hypothesized that if drivers drive on sharper curves while performing secondary tasks that they will have less unsafe and unstable eye movements for driving (H3).

Furthermore, from many previous studies of driver distraction, lane-keeping performance (e.g., minimizing standard deviations of lane position and steering wheel angle) decreases when drivers engage in a secondary task, particularly requiring higher visual attention to the task, even on straight roads (e.g., Reed & Green, 1999). Hurwitz and Wheatley (2002) found that the extent of lane-keeping variation was likely to be larger on curved roads compared with straight roads.

Several studies have focused on the curve's degree of sharpness: Wooldridge, Fitzpatrick, Koppa, and Bauer (2000) found that driver's visual demand was significantly increased while driving on curves with sharper radii. Tsimhoni and Green (2001) also found a similar result, visual demand increased when driving on sharper curves, when conducting a driving simulation study with voluntary occlusion techniques. In addition, higher visual demand was more likely to affect poorer driving performance. In the current study, I aimed to identify differences in driving performance and visual attention between different curvatures in multi-task settings. Thus, it is hypothesized that if drivers drive on sharper curves while performing secondary tasks they will have higher variability in lane-keeping performance (H4).

4.3. Method

4.3.1. Participants

A total of 24 licensed drivers participated in the experiment: 16 males and 8 females ranging in age between 19 and 31 years ($M = 22.6$, $SD = 3.53$). They reported that they drove more than 5,000 miles (average driving mileage = 24,480 miles) per year and they had adequate vision/hearing for driving. Participants were paid for their time with $15 hourly rate in cash.

4.3.2. Apparatus

A fixed-base driving simulator including a 24-inch LCD monitor (HP ZR24w) and a Logitech G27 RT racing wheel set (a force-feedback steering wheel, brake pedal, and accelerator) was used, as shown in Figure 4.1. A resolution of 1600 × 1200 (pixels) was used to generate the graphics of the driving scenarios on the monitor. The field of view was estimated approximately 38º × 25º, given that the display monitor area was 50 × 32 (H × V; cm²) and the viewing distance was 70-75 cm depending on the participants. An eye-tracking device (Gazepoint GP3; gazept.com)

was installed between the monitor and the steering wheel and used to collect subject's eye movement during the experiment. A touchscreen device (Samsung Galaxy Note 4; 2560 × 1440; 5.7-inch Quad HD Super AMOLED) was mounted on the right-hand side, 12-inch away from the center of the steering wheel to allow subjects to perform a secondary (stimulus-response) task while driving. MIT AgeLab NBack App (see Reimer, Gulash, Mehler, Foley, Arredondo, & Waldmann, 2014; required Android 4.0 and up) was installed on the touchscreen device. The illuminance of experimental space was 592 lux.

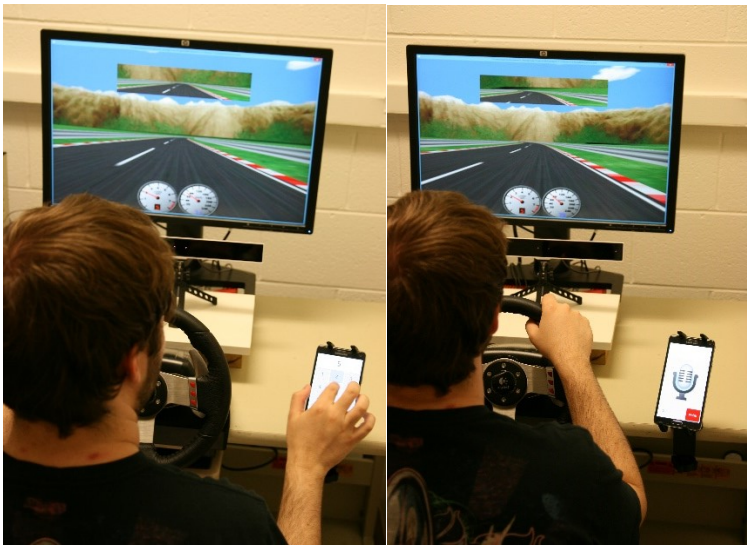


Figure 4.1 – Experimental setup of driving and stimulus-response tasks
(Visual-Manual task on left; Auditory-Speech task on right)

4.3.2.1. Driving scenarios

Eight driving scenarios were developed using The Open Racing Car Simulator (TORCS; torcs.sourceforge.net). Each scenario included curved roads with four levels of curvature radius (R = 100, 200, 400 and 800 meters; all curve deflection angles = 180 degrees), transition straight roads (300 m) between each curved road, and starting and ending roads (500 m for each). The curves were designed equally in both left and right directions. The eight scenarios including the

different order of four levels of the curvature radius and two directions were randomly given to the subjects; they were unaware which road they will drive.

Rural highway roads with two lanes in one direction were used and had a flat surface and only horizontal curves, no vertical curves. The width of each lane was 5 meters. There were approximately 4-meter-wide shoulders on each side of the roads. The average driving duration for each scenario was 4.5 minutes with an average speed of 50 - 60 km/h. The simulated vehicle was a mid-size sedan (Peugeot 406). A dashboard including a speedometer and a tachometer was displayed at the center-bottom of the screen. A rear-view mirror was presented at the center-top of the screen.

4.3.2.2. Stimulus-response task

To implement a driving distraction task (or a secondary task) in this study, a simple stimulus-response task (S-R task) with an n-back application software (i.e., MIT AgeLab NBack App) was used (see Figure 4.2). The n-back task (also called Delayed Digit Recall Task) was originally developed as a method to assess individual's working memory by perceiving a sequence of stimuli and repeating the one from n-steps earlier in the sequence. In this experiment, only the 0-back (the easiest level) was used to mainly focus on comparing the effect between four different stimulus-response types on driving performance, rather than cognitive demands from different types of n-back level. That is, the participants were asked to repeat aloud (i.e., speech response) or click the button (i.e., manual response) corresponding to the number that was 'just' presented on the screen (i.e., visual stimulus) or heard from the speaker (i.e., auditory stimulus). They were asked to respond to each stimulus as accurately as possible, but to prioritize the driving performance whenever they feel they need to improve driving safety.

The task included 2 different stimulus types (visual or auditory) and 2 different response types (manual or speech). The visual stimuli were composed of 10 Arabic numerals (0 to 9) and the auditory stimuli were the voice sounds of the 10 numerals. The vertical and horizontal size of the visual stimuli was 80 × 50 mm and the average contrast ratio (stimulus luminance/background luminance = 35/40) was .875. The average sound level of auditory stimuli was 75 dB, and was confirmed to be clearly audible by all subjects. Each stimulus was presented every 2.25 sec with a .75 sec in between each stimulus. One hundred visual or auditory stimuli were presented for 5 minutes: 3 minutes and 2 minutes were given separately because the maximum run duration for the software was 3 minutes. The order of separated stimuli duration was balanced across subjects. An experimenter gave the first stimulus of each duration (3 or 2 minutes) to the subject by pressing the 'Begin' button next to the subject so the stimulus-response task duration (5 minutes) sufficiently can cover the whole 4.5-minute driving task. The arrangement of the number buttons on the touchscreen was like a commercial telephone keypad and it was fixed during the whole experiment.

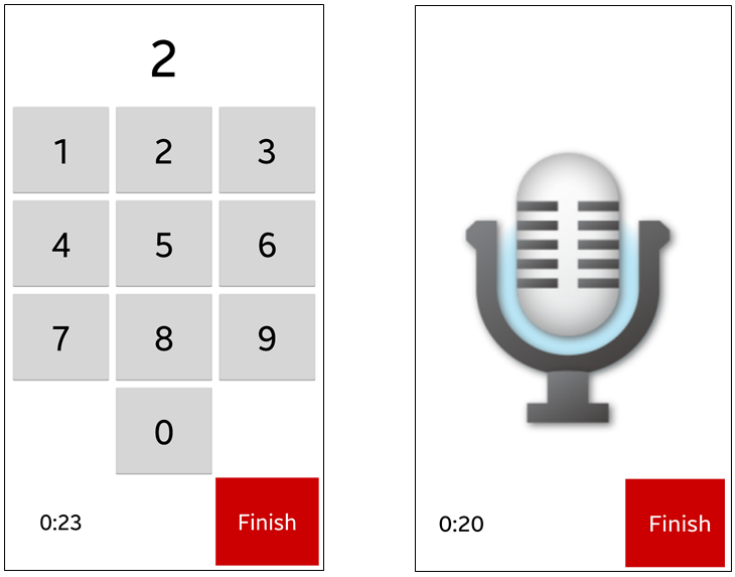


Figure 4.2 – A screen capture of the N-Back app (Reimer et al., 2014)
(Visual-Manual task on left; Auditory-Speech task on right)

4.3.3. Experimental design

Two independent variables in this experiment were stimulus-response (distraction) task type (S-R Type) and driving road curvature (Road Curvature). Four levels of the stimulus-response task type were auditory-manual [A-M], auditory-speech [A-S], visual-manual [V-M], and visual-speech [V-S]. Two levels of the road curvature were the sharp curve (curve radii are 100 m and 200 m) and the moderate curve (400 m and 800 m). The classification of the two levels was determined, based on the previous curve road studies (e.g., Tsimhoni & Green, 2001; Calvi, 2015).

Twenty-four subjects performed four sessions in total. All sessions included all of the two levels of driving road curvature, whereas each session had each individual secondary task. The order of each level for both independent variables was balanced across subjects.

4.3.4. Procedures

Prior to arrival at the laboratory, each participant was asked to fill out a brief online survey about their demographic information (e.g., age, gender) and driving experience (e.g., mileages driven, years). At the laboratory, each participant was informed of the purposes and nature of the experiment and they completed a consent form. Before the experiment started, participants performed practice trials for 10-15 minutes to become familiar with both driving and secondary tasks. After comprehensive training was provided to the participants, the calibration for the eye tracking system was conducted. Each subject performed all four sessions of dual-task which each of them took 5 to 7 minutes. During the experiment, participants were instructed to prioritize the driving task performance over the secondary task performance. After each session, they completed a DALI (Driving Activity Load Index; Pauzié, 2008) survey and had a 5-minute break to avoid fatigue from each dual-task. At the end of all sessions, they completed a payment form. The whole experiment took approximately an hour.

4.3.5. Dependent measures

4.3.5.1. Eye movements

For eye-tracking analysis, it is essential to determine the appropriate eye-tracking metrics, such as the area of interest, defined as an experimenter-defined region of the display. A typical area of interest used in driving study includes outside view of vehicle or dashboard. Klauer et al. (2006) categorized three areas of eye-glance (called Ellipse 1 - 3) commonly used in driving research, based on the degrees of visual angle away from the center forward view. In this study eyes-on-road behavior is defined as the eye-glance in the Ellipse 1 (Left Forward, Right Forward, and Instrument Panel; 20 degrees or less away from center forward) and eyes-off-road behavior refers to the outside of the Ellipse 1.

Unsafe and unstable eye movements for driving were measured by the percentage of eyes-off-road glances (%EORG), the percentage of eyes-off-road time (%EORT), the standard deviation of lateral (or yaw) eye fixation (SDLEF), and the standard deviation of vertical (or pitch) eye fixation (SDVEF). All these eye movement measures are considered as a standard measure of driving safety.

EORG and EORT represent the eye fixation frequency and duration not directed to the road respectively. These measures have been commonly used as a surrogate measure of driver inattention and visual-cognitive distraction in previous studies (e.g., Sodhi, Reimer, & Llamazares, 2002; Liang & Lee, 2010). In general, the total eyes-off-road time (TEORT) measure has been used to assess the design of in-vehicle electronic devices, as the maximum acceptable TEORT is 2 seconds and cumulatively 12 seconds (NHTSA, 2013). However, since the driving distance and time varies with the road curvature (i.e., obviously longer chance of TEORT at the moderate curve than the sharp curve) in this study, EORG and EORT were able to be used for the only comparison

between four different S-R Types, not between road curvature levels. Therefore, instead, %EORG and %EORT were used in this study for both effects of S-R Types and Road Curvature, defined as how many portion of eyes-off-road behavior occurs based on all the possible eye fixation frequencies and durations during the task.

SDLEF represents how much a driver has lateral eye movement. Kountouriotis and Merat (2016) used this measure to determine whether drivers scan the pattern of the scene (if higher) or concentrate towards the road ahead (if lower). By the same token, SDVEF represents how much a driver looks at the scene in the vertical direction. Since this study involves only the horizontal curves, not vertical elevations, it is expected that the effect of vertical eye movement would be mainly associated with looking at the touch screen device (to conduct secondary tasks) and instrumental panel (to check their driving speed).

For SDLEF and SDVEF measures, eye fixation's coordinate system had to be clearly determined. The coordinate system was customized using the monitor screen size and resolution (1600 × 1200). The origin of the coordinate system was set by the center of the monitor positioned the center of the road from the driver's perspective. For example, the four edges of the monitor were set as (1/2, 6/16), (-1/2, 6/16), (-1/2, -6/16), and (1/2, 6/16), clockwise from the right-upper corner. In other words, SDLEF and SDVEF values indicate the percentage of the monitor screen's horizontal (51.5 cm) and vertical (32 cm) size, respectively.

4.3.5.2. Variability in lane-keeping performance

Variability in lane-keeping performance was measured by standard deviation of lateral position (SDLP) and standard deviation of steering wheel angle (SDSWA). Lateral position indicates the distance gap between the center of the vehicle and the center of the driving lane. The steering angle refers to the angle between and the steered wheel direction and the front of the

vehicle. SDLP and SDSWA are a surrogate measure to evaluate the driver's lateral control and steering behavior used in previous driving studies (e.g., Jamson, Westernman, Hockey, & Carsten, 2004; Boyle, Tippin, Paul, & Rizzo, 2008; Dijksterhuis, Brookhuis, & De Waard, 2011; He, McCarley, & Krammer, 2014).

4.3.5.3. Subjective workload

Subjective workload was measured by six driving activity load measures with a scale ranging from 1 (lowest) to 7 (highest): effort of attention (i.e., attention level required by the driving task), visual demand (i.e., visual demand necessary for the driving task), auditory demand (i.e., auditory demand necessary for the driving task), temporal demand (i.e., specific constraint due to timing demand when running the driving task), interference (i.e., possible disturbance when running the driving task simultaneously with the secondary tasks), and situational stress (i.e., constraints/stress level while conducting driving task) (Pauzié, 2008). The subjective workload was measured only across the four S-R types, not the road curvature, because it was expected that participants are less likely to recognize the different road curvatures by each radius (while they could differentiate whether the curves are sharp or moderate) than different S-R types during the experiment.

4.3.6. Data Analysis

Most dependent variables except the subjective workload were analyzed using the repeated-measures two-way ANOVA 4 (*S-R Type*: A-M, A-S, V-M, and V-S) × 2 (*Road Curvature*: Sharp and Moderate) as within-subject factors in SPSS version 23. The subjective workload was analyzed by one-way ANOVA with four S-R Types. Prior to performing the analyses of variance, all dependent variables were examined to verify whether they have a normal distribution using Curran's criteria of skewness (< 2) and kurtosis (< 7) (Curran, West, & Finch 1996). Also, the

Levene's test for homogeneity of variance was used to examine the assumption of homogeneity of variance across groups. Effect sizes were obtained by partial eta-squared (η_p^2) because two independent variables were used in this study (Cohen, 1973; Brown, 2008); eta-squared (η^2) was used for subjective workloads. Additionally, post hoc tests were performed using Tukey's honest significant difference (HSD) to examine significant differences among each level of independent variables, especially for the S-R Type including more than two levels. A significant level was set at $\alpha = .05$.

4.4. Results

4.4.1. Descriptive statistics

The number of datasets collected with the frequency of 60Hz from the systems was 384 (24 subjects × 4 levels of S-R Type × 2 levels of Road Curvature × 2 replications): driving data (i.e., lateral positions and steering wheel angles) and eye movement data (i.e., eye fixation X-Y coordinates and durations of each fixation). However, since there were several missing data points due to procedural and equipment malfunction, 240 eye movement and 381 driving datasets were analyzed. Table 4.1 shows the summary of datasets by S-R Type and Road Curvature.

Table 4.1 – Descriptive statistics by S-R Type and Road Curvature

		S-R Type				Road Curvature	
		A-M	A-S	V-M	V-S	Sharp	Moderate
EORG (%)							
	Mean	25.5	.97	43.8	36.0	25.4	27.8
	(SD)	(6.26)	(2.85)	(7.56)	(7.71)	(17.3)	(17.7)
EORT (%)							
	Mean	16.1	.91	34.8	27.9	19.1	20.9
	(SD)	(5.10)	(3.12)	(8.31)	(8.49)	(14.4)	(14.7)
SDLEF (lateral size of screen = 1)							
	Mean	.49	.06	.57	.52	.41	.41
	(SD)	(.14)	(.02)	(.17)	(.11)	(.24)	(.24)
SDVEF (vertical size of screen = 1)							
	Mean	.63	015	.67	.58	.50	.51
	(SD)	(.10)	(.04)	(.07)	(.09)	(.23)	(.22)
SDLP (m)							
	Mean	.37	.30	.42	.34	.39	.32
	(SD)	(.16)	(.13)	(.19)	(.14)	(.19)	(.13)
SDSWA (degree)							
	Mean	.35	.22	.42	.30	.42	.22
	(SD)	(.22)	(017)	(.27)	(.19)	(.25)	(.13)
Workload - overall (1-7)							
	Mean	3.89	2.60	4.74	3.90	-	-
	(SD)	(1.03)	(.94)	(1.02)	(1.00)	(-)	(-)

4.4.2. Eye movements

4.4.2.1. Eye-glance pattern

Figure 4.3 represents a subject's eye-glance pattern depending on the S-R Type and Road Curvature. There was a tendency for the higher frequencies of eye fixation on the monitor screen (i.e., a 16:12 red square) when the subject drove on the sharp curve than on the moderate curve. Also, the eye fixations were mostly on the monitor screen during A-S task, but off the monitor screen during the other S-R task types.

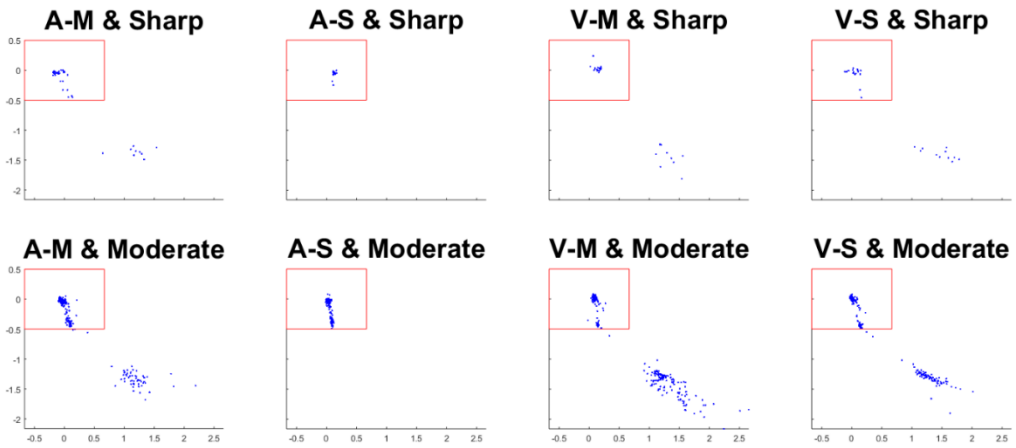


Figure 4.3 – A typical eye fixation pattern by S-R Type and Road Curvature
(Red squares indicate the monitor screen including Ellipse 1)

4.4.2.2. Percentage of eyes-off-road glances (%EORG)

4 × 2 ANOVA was run for the %EORG, which showed significant main effects of S-R Type (F (3, 240) = 530.38, p <.001, η_p^2 = .873) and Road Curvature (F (1, 240) = 7.99, p = .005, η_p^2 = .033); but there was no significant interaction between two factors. Post hoc test showed quite clear significant differences among the four S-R types: V-M > V-S > A-M > A-S (all p < .001). Not surprisingly, as shown in Figure 4.4, the V-M type (mean ($\bar{x}$) = 44.13, standard error of the mean (SEM) = .81) of distraction produced the highest %EORG, whereas the A-S type ($\bar{x}$ = .99, SEM = .81) activity made the lowest %EORG. The V-S type ($\bar{x}$ = 35.91, SEM = .84) made

higher percentage of eyes-off-road glances than A-M type ($\bar{x}$ = 25.62, *SEM* = .81). This finding is consistent with the first hypothesis (H1). That is, more frequent eyes-off-road behaviors (i.e., higher %EORG) were recorded on V-M, V-S, and A-M types, than A-S type.

The ANOVA showed a statistically significant difference on Road Curvature: moderate level of Road Curvature ($\bar{x}$ = 27.82, *SEM* = .62) resulted in higher %EORG than the sharp level of Road Curvature ($\bar{x}$ = 25.51, *SEM* = .53), as shown in Figure 4.5. This result indicates that drivers concentrated their visual attention on the road at the sharp curve than the moderate curve to maintain the driving safety. In other words, participants looked at the forward road 'more frequently' when driving on the sharp curve than the moderate curve. This result supports the third hypothesis (H3).

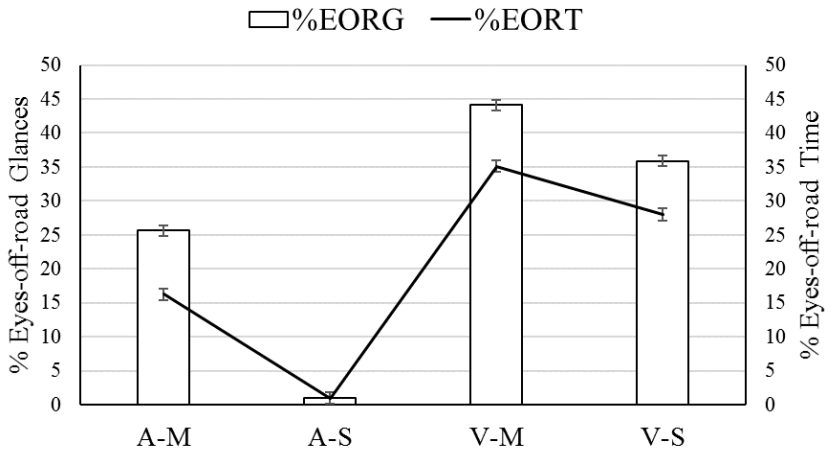


Figure 4.4 – Main effect of S-R Type on %EORG and %EORT (Error bar = *SEM*)

4.4.2.3. Percentage of eyes-off-road time (%EORT)

As shown Figure 4.4, ANOVA for %EORT revealed significant main effects for S-R Type (F (3, 240) = 305.54, $p < .001$, $\eta_p^2 = .798$) and Road Curvature (F (1, 240) = 4.97, $p = 0027$, $\eta_p^2 = .021$) but no interaction effect was found. Pearson correlation between %EORT and %EORG are quite highly correlated at 0.95. Post hoc test found that, similar to the result of %EORG, %EORT was significantly different (all $p < .001$) in this sequence: V-M (35.1 %) > V-S (28.0 %) > A-M (16.3 %) > A-S (.9 %). This finding is also consistent with the first hypothesis (H1). In other words, longer eyes-off-road behaviors (i.e., higher %EORT) were recorded on V-M, V-S, and A-M types than A-S type.

For the S-R Type comparison of frequency recording more than 12 seconds of TEORT (based upon the NHTSA guideline), the difference was in this sequence: V-M (24 times) > V-S (13 times) > A-M (12 times) > A-S (none). In other words, there was no big difference between V-S and A-M type.

The comparison for Road Curvature showed that %EORT was higher when the radius was moderate ($\bar{x} = 21.02$, $SEM = .66$) than sharp ($\bar{x} = 19.10$, $SEM = .56$), as shown in Figure 4.5. It was found that participants driving on a sharper curve looked at the forward road ‘longer’ than on the moderate curve, supporting the third hypothesis (H3).

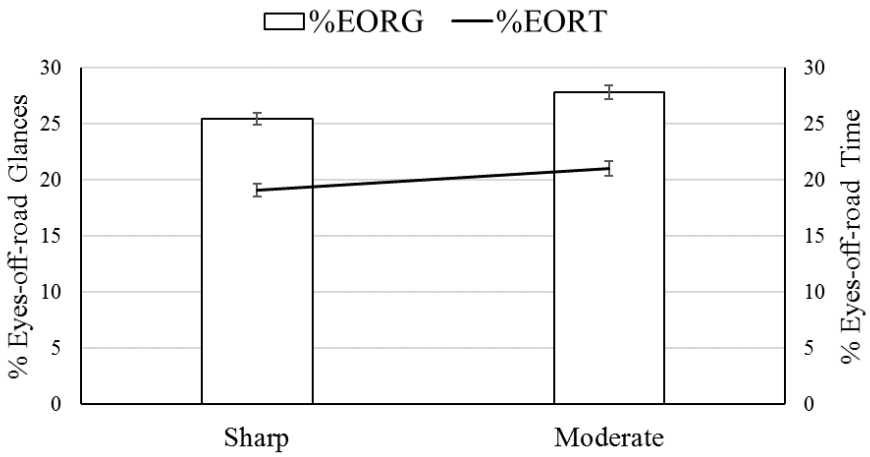


Figure 4.5 – Main effect of Road Curvature on %EORG and %EORT (Error bar = *SEM*)

4.4.2.4. Standard deviation of lateral eye fixation (SDLEF)

A significant main effect of S-R Type (F (3, 240) = 203.37, $p < .001$, $\eta_p^2 = .725$) on SDLEF was revealed, but the main effect of Road Curvature and the interaction effect were not found. V-M type (56.8 % of the lateral size of monitor screen), V-S (51.7 %), and A-M type (48.6 %) produced significantly higher SDLEF than A-S type (5.70 %), whereas V-M type caused higher SDLEF than A-M type, shown in Figure 4.6. However, significant differences on the V-S and A-M types ($p = .532$), as well as V-S and V-M types, were not found ($p = .114$).

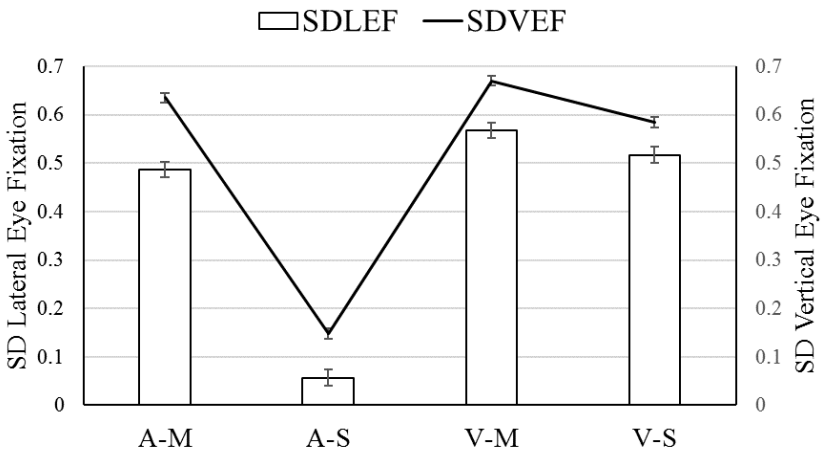


Figure 4.6 – Main effect of S-R Type on SDLEF and SDVEF (Error bar = *SEM*)

4.4.2.5. Standard deviation of vertical eye fixation (SDVEF)

Similar to the result of the SDLEF measure, only the main effect of S-R Type (F (3, 240) = 562.79, $p < .001$, $\eta_p^2 = .879$) was significant; no main effect of Road Curvature and interaction was found. However, in addition to the difference between V-M (67.0 % of the vertical size of monitor screen) and A-S (14.8 %) types, a pairwise comparison revealed a significant difference between A-M and V-S types, unlike the results of the SDLEF analysis. It showed that the value of SDVEF was significantly higher at the A-M Type (63.5 %) than the V-S Type (58.4 %) ($p = .003$), shown in Figure 4.6. This finding indicates that the vertical eye movement for looking at the stimulus on the touchscreen device had less variation than looking at the button to press (as a response) on the device. This result is a bit obvious because the location of the stimulus was closer to the road center than the location of response button.

4.4.3. Variability in lane-keeping performance

4.4.3.1. Standard deviation of lateral position (SDLP)

ANOVA for the SDLP revealed significant main effects of both S-R Type (F (3, 381) = 10.52, $p < .001$, $\eta_p^2 = .078$) and Road Curvature (F (1, 381) = 19.71, $p < .001$, $\eta_p^2 = .050$); but no interaction reached significance.

Pairwise comparisons, using a Tukey HSD procedure, revealed three statistically significant differences (all $p < .05$): V-M type ($\bar{x} = .42$, $SEM = .02$) produced higher SDLP than V-S type ($\bar{x} = .34$, $SEM = .02$) and A-S type ($\bar{x} = .30$, $SEM = .02$); SDLP for A-M type ($\bar{x} = .37$, $SEM = .02$) was higher than for A-S type. However, the comparison regarding which type of S-R modality between A-M and V-S types caused higher SDLP was not available. This result does not fully agree with the second hypothesis (H2). That is, the secondary task type that caused the least

visual demand (i.e., A-S type) resulted in the lowest variability in lane-keeping performance (i.e., lowest SDLP), but other comparisons were not available.

With regard to the comparison among the two Road Curvatures, the SDLP recorded along the sharp road curvature ($\bar{x}$ = .39, *SEM* = .01) was significantly higher than that on the moderate road curvature ($\bar{x}$ = .32, *SEM* = .01), shown in Figure 4.7. This finding was in agreement with the fourth hypothesis (H4).

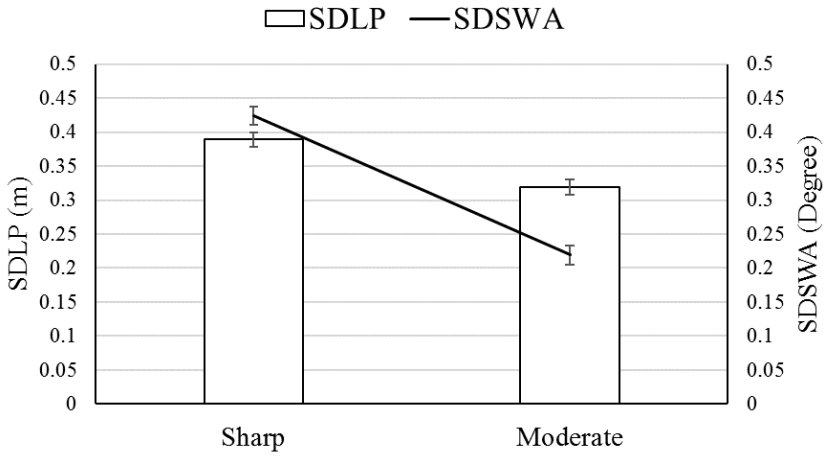


Figure 4.7 – Main effect of Road Curvature on SDLP and SDSWA (Error bar = *SEM*)

4.4.3.2. Standard deviation of steering wheel angle (SDSWA)

ANOVA showed significant main effects of and S-R Type (F (3, 381) = 17.85, $p < .001$, $\eta_p^2 = .126$) and Road Curvature (F (1, 381) = 110.77, $p < .001$, $\eta_p^2 = .229$) on the SDSWA, but no significant interaction effect was found. As shown Figure 4.7, it was revealed that SDSWA when subjects drove on the sharp curve ($\bar{x}$ = .42, *SEM* = .01) was significantly higher than when driving on the moderate curve ($\bar{x}$ = .22, *SEM* = .01). The fourth hypothesis was supported by this result, based on the result that variability in lane-keeping performance with secondary task increased (i.e., higher SDSWA) at the sharper curve.

Pairwise comparison test for S-R types found four significant differences: not surprisingly, V-M type ($\bar{x}$ = .42, *SEM* = .02) produced more SDSWA than A-S type ($\bar{x}$ = .222, *SEM* = .019) and V-S type ($\bar{x}$ = .30, *SEM* = .02). The A-M type ($\bar{x}$ = .35, *SEM* = .02) and V-S type produced more SDSWA than A-S type. However, no significant difference of SDSWA between A-M and V-S types was found. Similar to the SDLP analysis, this finding partially supports the second hypothesis (H2).

4.4.4. Subjective workload

The effect of S-R Type on driving activity load (DALI overall rating) was found to be significant (F (3, 92) = 18.92, $p < .001$, $\eta^2 = .382$). Pairwise comparisons revealed that the driving activity load was significantly different among the four S-R types (all $p < .001$), except between A-M and V-S types ($p > .05$). Although the six specific categories in the DALI questionnaire had similar outcomes, compared to the overall rating, the effect of S-R type in the 'Visual demand' category was revealed to be the most distinct difference among the four S-R types as shown in Figure 4.8. Especially, it was found that V-S type ($\bar{x}$ = 5.13, *SEM* = .25) task caused higher visual demand than A-M type ($\bar{x}$ = 4.17, *SEM* = .25) task (p = .038).

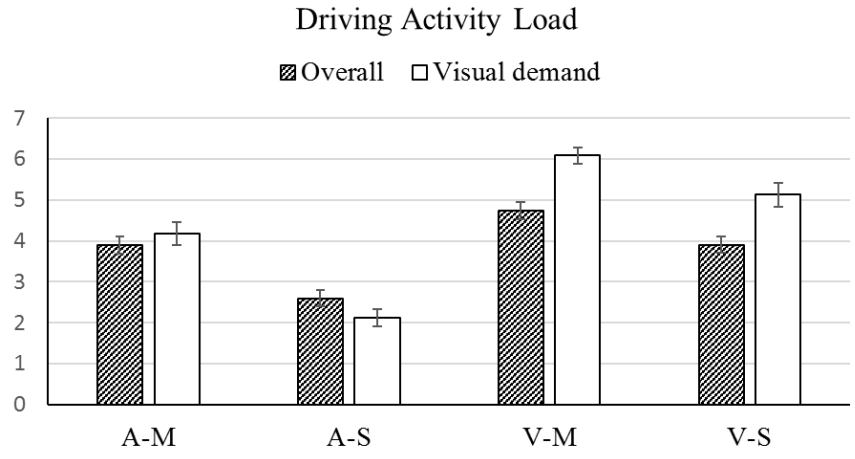


Figure 4.8 – Main effect of S-R Type on overall rating and visual demand (Error bars = *SEM*)

4.5. Discussion and conclusions

I investigated the effects of secondary-task modality as well as the degree of horizontal road curvature on drivers' visual attention patterns and their lane-keeping performances. Although no significant interaction was shown between the road curvature and the task modality, significant effects for two factors were clearly seen. Participants driving on more sharply curved roads took longer glances and more frequent glances at the road ahead (i.e., lower %EORG and %EORT), but interestingly, their lane-keeping performance was unstable (i.e., higher SDLP and SDSWA). As shown in the results of the secondary-task performance (i.e., a greater percentage of correct responses on moderate curves than sharp curves), participants prioritizing driving performance tended to focus more on safe driving on the sharper curves. For comparison of task modalities, the S-R types using a visual stimulus or a manual response (V-M, V-S, and A-M types) recorded poorer lane-keeping performance and more eyes-off-road moments than the S-R type using auditory stimulus and speech response (A-S type).

However, it was noted that it was hard to compare the performances of auditory-manual and visual-speech types. In fact, the comparison of performance between A-M and V-S types varied depending on the type of measures used. In other words, no significant difference between A-M and V-S types on stable driving control was found, whereas there was a significant difference in their effect on eyes-on-road behavior and subjective rating of visual demand: more eyes-off-road moments (i.e., higher %EORG and %EORT) were produced by V-S types than A-M types. Given that no, or less visual attention is generally required for auditory-stimulus (of A-M types) and speech-response (of V-S types), this finding showed that eyes-off-road moments were associated with visual attention induced by visual stimuli rather than manual responses. Participants actually reported that they felt higher visual demand from V-S types compared to A-

M types, as shown in Figure 4.8. A possible explanation for this finding might be top-down (expectation) mental processing, participants were not able to anticipate what visual stimuli would be presented in the experiment because Arabic numerals were randomly presented, whereas, they were able to anticipate which buttons to press due to the button's fixed positions, especially when they become acclimated to the task. As a result, the participants probably paid more visual attention to the display while performing V-S type tasks than A-M type tasks. Thus, the V-S modality elicited more eyes-off-road moments than the A-M modality.

There were some limitations in the current study's experimental setup. Firstly, in terms of a participant's perception of a curving road, Zakowska (1999) argued that perception of a static presentation is less sensitive to geometric curve properties than for a dynamic presentation. A two-dimension screen used in the current study probably affected the driver's perceptions; the driver's perception of the road curvature was likely less than its actual curvature. Secondly, since a road with larger curvature radius has a longer distance (i.e., exactly times of curvature radius because the curve deflection angles were 180 degrees for all curves), participants might have felt more fatigued when they drove on the moderate curves, which might have affected their performance of both the driving and secondary tasks. Thirdly, the narrow field of view used in the current study might affect participants' decisions for curve negotiation (Land & Horwood, 1995). In a limited field of view, the larger curves may have been less visible and peripheral cues may have been absent. Lastly, the selected curve radii in the driving simulation might not have necessarily reflected real-world conditions. For example, the Highway Design Manual (2012) recommends that a minimum speed for driving on a 100-meter radius curve in California is approximately 40 km/h. Given that the participants in this study drove on average 50–60 km/h, the 100-meter radius curve might have been too challenging to drive through even without any secondary tasks. It would

have been better to have had a baseline condition that participants drove to evaluate the road curvature selection, even though the authors confirmed that there was no issue for participants driving on sharp curves during practice trials.

In spite of these limitations, the results from the current study still provide useful guidelines with respect to use of secondary-task modalities for safe driving on curved roads. In other words, it would be better for drivers to use the methods using auditory stimuli or speech inputs, which require less visual demand, and to avoid methods using visual stimuli or manual inputs when they operate in-vehicle systems. That is based on the findings that A-S modality type tasks had the highest scores for lane-keeping performances and eyes-on-road behavior out of the four modality types.

However, it should be noted that there are several drawbacks to using A-S type of modalities or functions. First, the accuracy of speech interactions can be reduced by internal factors. Benzeghiba et al. (2007) argued that foreign and regional accents, rate of speech, and emotional states can cause higher variabilities for speech recognition. Second, the accuracy of speech recognition systems can be affected by external factors; for example, Moreno (1996) found that the accuracy of speech control systems decreases, especially when they are operated in noisy environments. Third, even though the use of auditory-speech interactions are known as less distracting methods compared to the use of visual-manual interactions while driving, mental workload still exists. Mental workload is higher when speech interactions are used while driving compared to a single driving task (Lo & Green, 2013; Wu, Chang, Boyle, & Jennes, 2015). Moreover, Wu et al. (2015) argued that task-load ratings while using speech systems varied with the systems used, such as portable voice control devices (e.g., smartphones) and built-in in-vehicle voice control systems. Because of drawbacks of speech recognition systems, drivers may need to

repeat voice commands because the voice commands are not correctly recognized by the systems. Sometimes, perceiving the auditory stimuli or voice announcements from the systems can fail due to interruption from surrounding conditions, such as ongoing dialogues with passengers and listening to music.

Further investigations are planned. First, I will examine driver cognitive demand while performing secondary tasks on curved roads (1) by measuring eye fixations only within the area of the road display and (2) by adding other n-back levels of the secondary tasks (i.e., 1-back or 2-back). Through the first approach, I propose analyzing whether different S-R modalities change patterns of road scanning independent of head-down glances based on previous studies that have shown a tendency for cognitively distracted drivers to narrow their scanning range (e.g., Recarter & Nunes, 2003; Strayer, Drews, & Johnston, 2003; Kaber, Liang, Zhang, Rogers, & Gangakhedkar, 2012). Through the second approach, I propose to analyze driver behavior while they are performing A-S tasks in more realistic conditions (e.g., talking on the phone) that require central executive processes using working memory.

Second, I propose developing a cognitive model to predict drivers' lateral-control performances and eye-glance patterns depending on the conditions of both secondary-task types and road curvatures. Some preliminary modeling studies have been conducted by the authors using the Queueing Network Cognitive Architecture (Liu, Feyen, & Tsimhoni, 2006). While current models cover each condition separately (i.e., secondary-task type in Jeong & Liu, 2017; road curvature in Jeong, Feng, & Liu, 2017), I propose dealing with both conditions in a single model. It is expected that this model will help understand the mechanisms of drivers' behaviors with different modality tasks on various horizontal curves.

References

Angell, L. S., Auflick, J., Austria, P. A., Kochhar, D. S., Tijerina, L., Biever, W., ... & Kiger, S. (2006). *Driver workload metrics task 2 final report* (No. HS-810 635) Washington, DC: U.S. Department of Transportation National Highway Traffic Safety Administration.

Benedetto, A., Calvi, A., & D'Amico, F. (2012). Effects of mobile telephone tasks on driving performance: a driving simulator study. *Advances in Transportation Studies*, *26*, 29-44.

Benzeghiba, M., De Mori, R., Deroo, O., Dupont, S., Erbes, T., Jouvet, D., ... & Rose, R. (2007). Automatic speech recognition and speech variability: A review. *Speech Communication*, *49*(10), 763-786.

Boyle, L. N., Tippin, J., Paul, A., & Rizzo, M. (2008). Driver performance in the moments surrounding a microsleep. *Transportation Research Part F: Traffic Psychology and Behaviour*, *11*(2), 126-136.

Brown, J. D. (2008). Effect size and eta squared. *JALT Testing & Evaluation SIG Newsletter*, *12*(2), 38-43.

Caliendo, C., Guida, M., & Parisi, A. (2007). A crash-prediction model for multilane roads. *Accident Analysis & Prevention*, *39*(4), 657-670.

Calvi, A. (2015). A study on driving performance along horizontal curves of rural roads. *Journal of Transportation Safety & Security*, *7*(3), 243-267.

Charlton, S. G. (2007). The role of attention in horizontal curves: A comparison of advance warning, delineation, and road marking treatments. *Accident Analysis & Prevention*, *39*(5), 873-885.

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement*, *33*(1), 107-112.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*(1), 16.

Dijksterhuis, C., Brookhuis, K. A., & De Waard, D. (2011). Effects of steering demand on lane keeping behaviour, self-reports, and physiology. A simulator study. *Accident Analysis & Prevention*, *43*(3), 1074-1081.

Dingus, T.A., Klauer, S.G., Neale, V.L., Petersen, A., Lee, S.E., Sudweeks, J., Perez, M.A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z.R., Jermeland, J., & Knipling, R.R. (2006). *The 100-Car Naturalistic Driving Study: Phase II – Results of the 100-Car Field Experiment*. Washington, DC: National Highway Traffic Safety Administration, USDOT. (Report No. DOT HS 810 593).

Fitzpatrick, K., Lord, D., & Park, B. J. (2010). Horizontal curve accident modification factor with consideration of driveway density on rural four-lane highways in texas. *Journal of Transportation Engineering*, *136*(9), 827-835.

Glennon, J. C., Neuman, T. R., & Leisch, J. E. (1985). *Safety and Operational Considerations for Design of Rural Highway Curves.* Report No. FHWA-RD-86-035, Federal Highway Administration, U.S. Department of Transportation.

He, J., McCarley, J. S., & Kramer, A. F. (2014). Lane keeping under cognitive load: Performance changes and mechanisms. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *56*(2), 414-426.

Hibberd, D. L., Jamson, S. L., & Carsten, O. M. (2013). Mitigating the effects of in-vehicle distractions through use of the Psychological Refractory Period paradigm. *Accident Analysis & Prevention*, *50*, 1096-1103.

Highway Design Manual. (2012). Sacramento, CA: California Department of Transportation. http://www.dot.ca.gov/design/manuals/hdm.html.

Horrey, W. J., Wickens, C. D., & Consalus, K. P. (2006). Modeling drivers' visual attention allocation while interacting with in-vehicle technologies. *Journal of Experimental Psychology: Applied*, *12*(2), 67.

Hurwitz, J. B., & Wheatley, D. J. (2002, September). Using driver performance measures to estimate workload. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 46, No. 22, pp. 1804-1808). Sage CA: Los Angeles, CA: SAGE publications.

Jamson, A. H., Westerman, S. J., Hockey, G. R. J., & Carsten, O. M. (2004). Speech-based e-mail and driver behavior: Effects of an in-vehicle message system interface. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(4), 625-639.

Janssen, C. P., & Brumby, D. P. (2010). Strategic adaptation to performance objectives in a dual-task setting. *Cognitive science*, *34*(8), 1548-1560.

Jeong, H., Feng, F., & Liu, Y. (2017). Computational Modeling of Driver Later Control on Curved Roads with Integration of Vehicle Dynamics and Reference Trajectory Tracking, In *Proceedings of the Ninth International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (pp.193-199).

Jeong, H., & Liu, Y. (2017). Modeling of Stimulus-Response Secondary Tasks with Different Modalities while Driving in a Computational Cognitive Architecture. In *Proceedings of the Ninth International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (pp. 58-63).

Johnston, I. R., (1982). Modifying driver behaviour on rural road curves: a review of recent research. *Proceedings of the 11th Australian Road Research Board (ARRB) Conference*, Vol. 11(4), pp. 115-124.

Kaber, D. B., Liang, Y., Zhang, Y., Rogers, M. L., & Gangakhedkar, S. (2012). Driver performance effects of simultaneous visual and cognitive distraction and adaptation behavior. *Transportation Research Part F: Traffic Psychology and Behaviour*, *15*(5), 491-501.

Klauer, S.G., Dingus, T.A., Neale, V.L., Sudweeks, J.D., & Ramsey, D.J., (2006). The Impact of Driver Inattention on Near-crash/Crash Risk: An Analysis Using the 100-car Naturalistic Driving Study Data (No. DOT HS 810 594). *Virginia Tech Transportation Institute*.

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

Kun, A. L., Brumby, D. P., & Medenica, Z. (2014, September). The musical road: Interacting with a portable music player in the city and on the highway. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 1-7). ACM.

Land, M., & Horwood, J. (1995). Which parts of the road guide steering?. *Nature*, *377*(6547), 339.

Liang, Y., & Lee, J. D., (2010). Combining cognitive and visual distraction: Less than the sum of its parts. *Accident Analysis & Prevention, 42*(3), 881-890.

Liu, Y. C., (2003). Effects of using head-up display in automobile context on attention demand and driving performance. *Displays, 24*(4), 157-165.

Liu, C., & Subramanian, R. (2009). Factors related to fatal single-vehicle run-off-road crashes (No. HS-811 232), Washington, DC, U.S. Department of Transportation.

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, V. E.-W., & Green, P. A. (2013). Development and Evaluation of Automotive Speech Interfaces: Useful Information from the Human Factors and the Related Literature. *International Journal of Vehicular Technology*, *2013*, 1–13.

McDonald, L.B., & Ellis, N.C., (1975). Driver workload for various turn radii and speeds. Transport. Res. Rec. 530, 8–30.

McGee, H. W. (2011). *Geometric Design Practices for Resurfacing, Restoration, and Rehabilitation* (Vol. 417). Transportation Research Board.

Milton, J., & Mannering, F., (1998). The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. *Transportation, 25*(4), 395-413.

Moreno, P. J., (1996). Speech recognition in noisy environments (Doctoral dissertation, Carnegie Mellon University).

National Highway Traffic Safety Administration. (NHTSA, 2013). *Visual-manual NHTSA Driver Distraction Guidelines for In-vehicle Electronic Devices* (No. NHTSA-2010-0053). Washington, DC: National Highway Traffic Safety Administration.

Pauzié, A. (2008). A method to assess the driver mental workload: The driving activity load index (DALI). *IET Intelligent Transport Systems*, *2*(4), 315-322.

Peng, Y., Boyle, L. N., & Hallmark, S. L., (2013). Driver's lane keeping ability with eyes off road: insights from a naturalistic study. *Accident Analysis & Prevention, 50*, 628-634.

Recarte, M. A., & Nunes, L. M. (2003). Mental workload while driving: effects on visual search, discrimination, and decision making. *Journal of Experimental Psychology: Applied*, *9*(2), 119.

Reed, M. P., & Green, P. A., (1999). Comparison of driving performance on-road and in a low-cost simulator using a concurrent telephone dialling task. *Ergonomics, 42*(8), 1015-1037.

Reimer, B., Gulash, C., Mehler, B., Foley, J. P., Arredondo, S., & Waldmann, A. (2014, September). The MIT AgeLab n-back: a multi-modal android application implementation. In *Adjunct Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 1-6). ACM.

Sodhi, M., Reimer, B., & Llamazares, I. (2002). Glance analysis of driver eye movements to evaluate distraction. *Behavior Research Methods*, *34*(4), 529-538.

Stelzel, C., & Schubert, T., (2011). Interference effects of stimulus–response modality pairings in dual tasks and their robustness. *Psychological Research, 75*(6), 476-490.

Strayer, D. L., Drews, F. A., & Johnston, W. A. (2003). Cell phone-induced failures of visual attention during simulated driving. *Journal of Experimental Psychology: Applied*, *9*(1), 23.

Tsimhoni, O., & Green, P., (2001). Visual demand of driving and the execution of display-intensive in-vehicle tasks. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 45, No. 23, pp. 1586-1590). SAGE Publications.

Vollrath, M., & Totzke, I. (2000, June). In-vehicle communication and driving: An attempt to overcome their interference. In *Driver Distraction Internet Forum sponsored by the United States Department of Transportation*.

Wickens, C. D., Sandry, D. L., & Vidulich, M. (1983). Compatibility and resource competition between modalities of input, central processing, and output. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 25*(2), 227-248.

Wickens, C. D., Vidulich, M., & Sandry-Garza, D. (1984). Principles of SCR compatibility with spatial and verbal tasks: The role of display-control location and voice-interactive display-control interfacing. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *26*(5), 533-543.

Wooldridge, M., Fitzpatrick, K., Koppa, R., & Bauer, K. (2000). Effects of horizontal curvature on driver visual demand. *Transportation Research Record: Journal of the Transportation Research Board*, (1737), 71-77.

Wu, J., Chang, C. C., Boyle, L. N., & Jenness, J. (2015). Impact of In-vehicle Voice Control Systems on Driver Distraction Insights From Contextual Interviews. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 59, No. 1, pp. 1583-1587). SAGE Publications.

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

Zakowska, L., (1999). Road curve evaluation based on road view perception study. *Transportation Research Record: Journal of the Transportation Research Board*, (1689), 68-72.

Zegeer, C., Stewart, R., Reinfurt, D., Council, F., Neuman, T., Hamilton, E., ..., &Hunter, W., (1990). Cost-effective geometric improvements for safety upgrading of horizontal curves, Publication FHWA-RD-90-074. Federal Highway Administration, U.S. Department of Transportation, Washington D.C.

Chapter 5.

Queuing Network Modeling of In-vehicle Secondary Task Performance

5.1. Introduction

While driving, drivers continue to interact with in-vehicle systems and their surrounding environment, by performing a variety of secondary tasks (e.g., tuning a radio) in addition to driving a vehicle. Most of the secondary tasks are composed of multiple stimuli and their corresponding responses. In other words, in-vehicle secondary tasks represent a type of stimulus-response task: once drivers perceive stimuli (or receive information) from the in-vehicle systems, they often respond to the systems. While the conventional in-vehicle secondary tasks were the visual-manual type (e.g., rotating a knob while looking at the current radio tuning frequency), more diverse types have become common, using a wide range of modalities. For example, recent in-vehicle systems allow drivers to hear a voice announcement from the electronic navigation systems and to say a voice-command to input the information to the systems.

Many experimental studies have investigated the effect of different types of stimulus-response tasks on driving performance (e.g., Angell et al., 2006; Shutko, Mayer, Laansoo, & Tijerina, 2009; Young, Hsieh, & Seaman, 2013; Reimer et al., 2014b). According to the literature, visual and auditory modalities are two of the most frequently used information presentation channels in the in-vehicle secondary tasks, whereas manual and speech (or verbal) input techniques are the most common responding methods. In general, one of the common findings is that driving performance during the visual-manual task was significantly different

from that during the auditory-speech task, such as showing higher steering wheel reversal rates and higher ratio of eyes-off-road time. However, few studies have examined the more diverse types of modalities, such as auditory-manual [A-M], auditory-speech [A-S], visual-manual [V-M], and visual-speech [V-S] stimulus-response tasks. Furthermore, there are few modeling studies for predicting driving behavior and workload during the secondary tasks, even though modeling studies enable systems designers to find solutions to usability issues at an early stage of system development, thereby reducing labor and time cost.

In this chapter, I report a computational model to predict eyes-off-road behavior and workload in performing four different types of stimulus-response tasks (i.e., A-M, A-S, V-M, and V-S), using the Queueing Network-Model Human Processor (QN-MHP; Liu, Feyen, & Tsimhoni, 2006) which enables the multitasking prediction as well as human-machine interface (HMI) evaluation. The model used the interface of the MIT AgeLab NBack App to evaluate simple stimulus-response tasks (see Reimer et al., 2014a for details) and was evaluated with human subject data from 24 participants.

5.2. Model development

5.2.1. Task Analysis

Task analyses were conducted for all the four stimulus-response tasks: A-M, A-S, V-M, and V-S, using NGOMSL (Natural Goals, Operators, Methods, and Selection rules Language)-style task description (Kieras, 1999). Table 5.1 shows the result of tasks analysis. In the NGOMSL task analyses, task components (TCs) were used to describe each step to accomplish the whole task. Each TC is made with a pre-determined operator that runs with one or multiple parameter(s). In Table 5.1, showing the result of the task analyses, "*Look-at*", "*Listen-to*",

"*Click-with-finger*" are the examples of the operator, whereas "<target type>", "<device id>", "<*x*, *y*>" are the examples of the parameter.

Table 5.1 – NGOMSL-style descriptions of the four stimulus-response tasks

[A-M task]

Goal: Listen to a number and press the number button
TC 1: *Listen-to* <target value>
TC 2: *Store* the <target value> on STM*
TC 3: *Retrieve* the <target value > from STM
TC 4: *Compare* <target value> to <expected value> If match, return result =1, else return result = 0
TC 5: *Decide* If result = 1, go to TC 6, else go to TC 1
TC 6: *Look-at* <target type> on <device id> at location <x_1,y_1>
TC 7: *Store* the <target value> on STM
TC 8: *Retrieve* the <target value > from STM
TC 9: *Compare* <target value> to <expected value> If match, return result =1, else return result = 0
TC 10: *Decide* If result = 1, go to TC 11, else go to TC 6
TC 11: *Look-at* <target type> on <device id> at location <x_1,y_1>
TC 12: *Store* the <target value> on STM
TC 13: *Retrieve* the <target value> from STM
TC 14: *Determine-hand-movement*
TC 15: *Reach-with-hand*
TC 16: *Look-at* <target value> on <device id> at location <x_1,y_1>
TC 17: *Store* the <target value> on STM
TC 18: *Retrieve* the <target value> from STM
TC 19: *Determine-finger-movement*
TC 20: *Click-with-finger*
TC 21: Return with goal accomplished

*STM = short-term-memory

[V-M task]

Goal: Look at a number and press the number button
TC 1: *Wait/Find a visual stimulus* and *Look-at* <target type> on <device id> at location <x_0,y_0>
TC 2: *Store* the <target value> on STM
TC 3: *Retrieve* the <target value> from STM
TC 4: *Compare* <target value> to <expected value> If match, return result =1, else return result = 0
TC 5: *Decide* If result = 1, go to TC 6, else go to TC 1
TC 6: *Look-at* <target type> on <device id> at location <x_1,y_1>
TC 7: *Store* the <target value> on STM
TC 8: *Retrieve* the <target value> from STM
TC 9: *Compare* <target value> to <expected value> If match, return result =1, else return result = 0
TC 10: *Decide* If result = 1, go to TC 11, else go to TC 6
TC 11: *Look-at* <target type> on <device id> at location <x_1,y_1>
TC 12: *Store* the <target value> on STM
TC 13: *Retrieve* the <target value> from STM
TC 14: *Determine-hand-movement*
TC 15: *Reach-with-hand*
TC 16: *Look-at* <target type> on <device id> at location <x_1,y_1>
TC 17: *Store* the <target value> on STM
TC 18: *Retrieve* the <target value> from STM
TC 19: *Determine-finger-movement*
TC 20: *Click-with-finger*
TC 21: Return with goal accomplished

[A-S task]

Goal: Listen to a number and say the number
TC 1: *Listen-to* <target value>
TC 2: *Store* the <target value> on STM
TC 3: *Retrieve* the <target value> from STM
TC 4: *Compare* <target value> to <expected value> If match, return result =1, else return result = 0
TC 5: *Decide* If result = 1, go to TC 6, else go to TC 1
TC 6: *Say <a number>*
TC 7: Return with goal accomplished

[V-S task]

Goal: Look at a number and say the number
TC 1: *Wait/Find a visual stimulus* and *Look-at* <target type> on <device id> at location <x_0,y_0>
TC 2: *Store* the <target value> on STM
TC 3: *Retrieve* the <target value> from STM
TC 4: *Compare* <target value> to <expected value> If match, return result =1, else return result = 0
TC 5: *Decide* If result = 1, go to TC 6, else go to TC 1
TC 6: *Say <a number>*
TC 7: Return with goal accomplished

5.2.2. Development of Operators

Because the recent QN-MHP-based models have only operators for the tasks using visual stimuli and/or manual responses (Feng, Liu, Chen, Filev, & To, 2014; Jeong & Liu, 2016), two new operators were needed and thus developed for auditory stimuli and/or speech responses: *Listen-to* and *Say*. For these operators, it was assumed that each syllable takes the same amount of time for both listening and saying. A pre-determined audio library module including simple syllable-separated words for Arabic numerals (e.g., ze-ro, one, …, sev-en, …, nine) was used to implement these operators. Here, I describe four major operators including the two new operators used for investigating the four stimulus-response tasks:

Look-at: This operator allows a human model to look at a specific location. The specific target location is set with three parameters: type of target (e.g., text or color), device id, and a target's two-dimensional coordinates on the device. Once the "Look-at" operator is activated at Server D, a long-term procedural memory server, it triggers a saccade motor action at Server W, a motor-elements server. Then Server W triggers the Eyes server so a saccade can be executed at the Eyes server. The saccade execution time is determined by an angular velocity (i.e., 4 msec / degree; Kieras & Meyer, 1997) and a visual angle (i.e., angle from the current location of visual attention to the target location). Once the saccade is completed at the Eyes server, an entity (or visual stimulus) of target enters into Server 1, a visual input server. Then the entity enters Servers 2 (Visual recognition) and 3 (Visual location) and makes the human model recognize the visual target and its location, respectively. Through Server 4, a visual integration server, the entity is transformed into the cognitive subnetwork.

Listen-to: This operator allows a human model to listen to a text-based content (e.g., a syllable, a word, and a sentence) from a source of sound, such as an in-vehicle speaker. Once the

"Listen-to" operator is activated at Server D, it triggers an auditory motor action at Server W. Then Server W triggers the Ears server so a listening action can be executed at the Ears server. The listening execution time is determined by an internal process time and an external process time (or transmission time). The internal process time is assumed from the perception time randomly assigned, ranging from 50 to 200 msec (Card, Moran, & Newell, 1983), whereas the external process time (or transmission time) is determined by the distance from the sound source to human model's ears and a sound speed (i.e., 343.2 m/s). Once the listening action is completed at the Ears server, an entity (or auditory stimulus) enters into Server 5, an auditory input server. Then the entity enters Servers 6 (Auditory recognition) and 7 (Auditory location) and makes the human model recognize the sound and its location, respectively. Through Server 8, an auditory integration server, the entity is transformed into the cognitive subnetwork.

Reach-with-hand / Click-with-finger: These operators initiate a reaching and a clicking action using the model's hand servers. Once these operators are activated at Server D, a motor entity is created in Server W with the motor type of "Reach-with-hand" / "Click-with-finger". These motor entities are then processed in Servers W, Y, Z, and the Right-hand or Left-hand servers. The hand servers make the hand be reached to the target (or the finger be clicked on the target), based on the estimation of how far/long the hand reaches the target (or the finger clicks on the target). The reaching execution time is determined by the general Fitts' law equation. According to Shannon formulation (MacKenzie, 1992), the movement time MT is:

$$MT = a + b \times \log_2\left(\frac{A}{W} + 1\right) \qquad (1)$$

, where *a* and *b* are empirical regression coefficients, varying in the environment, such as people and devices. *W* refers to the target's size, whereas *A* refers to the distance to the target. The clicking execution time is determined as 280 msec from the Keystroke Level Model (Card et al.,

1980). For the manual response in this study, it is assumed that the model uses a right hand. Also, it is assumed that the reaching distance is 300 mm, which closely resembles the actual distance from a steering wheel to the target on the device.

Say: This operator initiates a speech (verbal) response action using the model's mouth server. Once this operator is activated at Server D, a motor entity is created in Server W with the motor type of "Say". This motor entity is then processed in Servers W, Y, Z, and the Mouth server. The Mouth server makes the model say a text-based content (e.g., a syllable, a word, and a sentence) and the corresponding button on the device is clicked. The speech response's execution time is determined with John (1990)'s finding, 130 - 170 msec per syllable, depending on the practiced level. In the current study, 130 msec per syllable was used, assuming the highly practiced level.

5.2.3. Development of Digital Device Mockups

Using MATLAB Graphical User Interface Design Environment (GUIDE), digital device mockups of the NBack App were developed. Figure 5.1 shows the digital mockup of the V-M task, as an example. Figure 5.2-(a) shows the coordinates for a visual stimulus (i.e., (x_0,y_0)) and a button for the manual response (i.e., (x_1,y_1)). The flow of the process when V-M task is performed is shown in Figure 5.2-(b). Where the model looks at and clicks on the digital mockup are indicated by yellow-hatched and white-dotted squares, respectively.

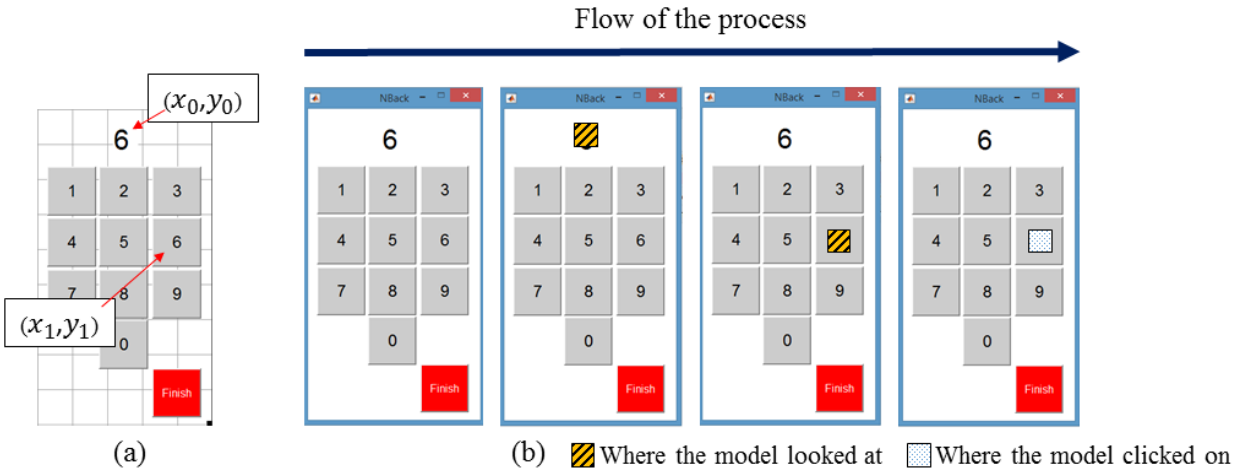


Figure 5.1 – (a) An example of digital mockup for the NBack App and (b) its usage for a visual-manual secondary task while driving

5.3. Model validation

Experimental data from 24 college students were used to evaluate the model. Participants (age: M = 22.6, SD = 3.53; 16 males and 8 females) were asked to perform the four stimulus-response tasks, using the 0-back task (the easiest level) from the NBack App software, while driving on simulated horizontal curves including multiple curvature levels (radii ranging from 100 to 800 meter). Since the purpose of the current study was to compare only the effects among the four different stimulus-response types, other levels of n-back tasks were not used. Either visual or auditory stimulus was presented for 2.25 s with a 0.75 s time gap between each stimulus. After the driving and secondary tasks, participants completed a Driving Activity Load Index (DALI; Pauzié, 2008) survey, as subjective workload measurements. The survey included six measures with a seven-level scale ranging from 1 (lowest) to 7 (highest), including effort of attention, visual demand, auditory demand, temporal demand, interference, and situational stress. The overall scores combining all the six measures were used to validate the model's workload outputs. Eye movement data were collected by Gazepoint GP3 at 60 Hz. In this study, eyes-off-

road time was defined as the duration of eye-glances in 20 degree or less away from the center forward (i.e., left and right forward, and instrument panel; Klauer, Dingus, Neale, Sudweeks, & Ramsey, 2006).

Ten model simulations were run, in which the eye location (recorded by the current location of visual attention) and the workload (estimated by server utilizations in the QN framework; normalized to 1-7 levels) data were collected every 50 msec. As shown in Figure 5.2, the model was able to generate results for both the percentage of eyes-off-road time (R^2 = 0.88, RMS = 4.95) and workload (R^2 = 0.99, RMS = 1.16) quite similar to the human subject data.

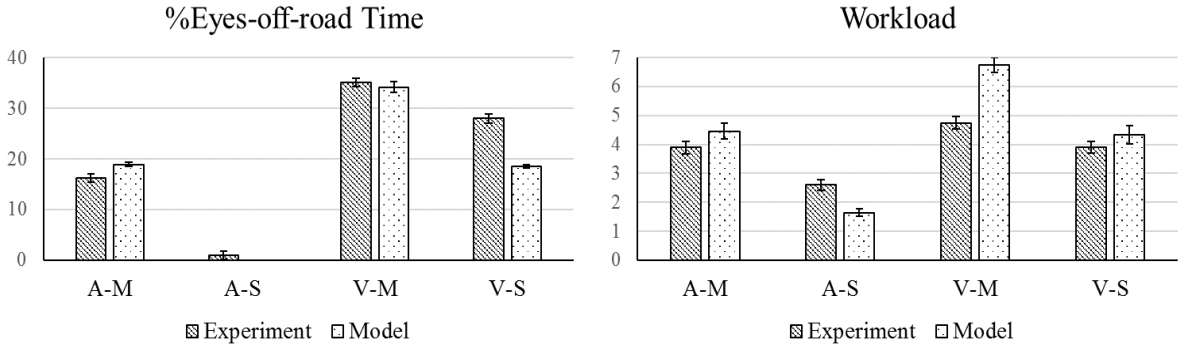


Figure 5.2 – Modeling results of %Eyes-off-road time and workload in comparison to human results

5.4. Conclusion and discussion

This chapter presented a computational model in the QN-MHP architecture for the four stimulus-response tasks of the combinations of two stimuli (i.e., auditory and visual) and two responses (i.e., manual and speech). In addition to developing a predictive model, I evaluated the model with empirical data from 24 subjects, and found very good validation results in the time ratio of eyes-off-road as well as workload (more than 85 % of R^2 for both outputs; less than 5% of RMS for Eyes-off-road time and less than 1.5 levels of RMS for workload).

To model the tasks using auditory stimuli and/or speech responses, two new auditory-related operators were developed and they were implemented with a pre-determined audio library module. Since the current module includes only Arabic numerals (i.e., 0 - 9) pronunciation and its syllable breakdown, further study aims to extend the types of voice commands actually used in the practical driving setting (e.g., 'say a command', 'increase the temperature') so that I can support HMI evaluations with a wider range of usability test for in-vehicle infotainment system developments.

References

Angell, L. S., Auflick, J., Austria, P. A., Kochhar, D. S., Tijerina, L., Biever, W., ... & Kiger, S. (2006). *Driver workload metrics task 2 final report* (No. HS-810 635) Washington, DC: U.S. Department of Transportation National Highway Traffic Safety Administration.

Card, S. K., Moran, T. P., & Newell, A. (1980). The keystroke-level model for user performance time with interactive systems. *Communications of the ACM*, 23(7), 396-410.

Feng, F., Liu, Y., Chen, Y., Filev, D., & To, C. (2014, September). Computer-Aided Usability Evaluation of In-Vehicle Infotainment Systems. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, No. 1, pp. 2285-2289). SAGE Publications.

Jeong, H., & Liu, Y. (2016, September). Computational Modeling of Finger Swipe Gestures on Touchscreen Application of Fitts' Law in 3D Space. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 60, No. 1, pp. 1721-1725). SAGE Publications.

John, B. E. (1990, March). Extensions of GOMS analyses to expert performance requiring perception of dynamic visual and auditory information. In *Proceedings of the SIGCHI conference on Human factors in Computing Systems* (pp. 107-116). ACM.

Kieras, D. E., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction*, *12*(4), 391-438.

Kieras, D. E. (1999). A guide to GOMS model usability evaluation using GOMSL and GLEAN3. *University of Michigan*, (313).

Klauer, S.G., Dingus, T.A., Neale, V.L., Sudweeks, J.D., & Ramsey, D.J., (2006). The Impact of Driver Inattention on Near-crash/Crash Risk: An Analysis Using the 100-car Naturalistic Driving Study Data (No. DOT HS 810 594). *Virginia Tech Transportation Institute*.

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.

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

Pauzié, A., 2008. A method to assess the driver mental workload: The driving activity load index (DALI). *IET Intelligent Transport Systems*, 2(4), 315-322.

Reimer, B., Gulash, C., Mehler, B., Foley, J. P., Arredondo, S., & Waldmann, A. (2014a, September). The MIT AgeLab n-back: a multi-modal android application

implementation. In *Adjunct Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 1-6). ACM.

Reimer, B., Mehler, B., Dobres, J., McAnulty, H., Mehler, A., Munger, D., & Rumpold, A. (2014b, September). Effects of an 'Expert Mode' Voice Command System on Task Performance, Glance Behavior & Driver Physiology. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 1-9). ACM.

Shutko, J., Mayer, K., Laansoo, E., & Tijerina, L. (2009). *Driver workload effects of cell phone, music player, and text messaging tasks with the Ford SYNC voice interface versus handheld visual-manual interfaces* (No. 2009-01-0786). SAE Technical Paper.

Young, R. A., Hsieh, L., & Seaman, S. (2013, June). The tactile detection response task: preliminary validation for measuring the attentional effects of cognitive load. In *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (pp. 71-77).

Chapter 6.

An Experimental Study on the Effects of Road Geometry and Lead Vehicle on Driving Performance

6.1. Introduction

According to the Transportation Research Board reports (Torbic et al., 2004; Federal Highway Administration, 2006), nearly 25 percent of fatal crashes occur along horizontal curved roads, predominantly on two-lane local highways. Moreover, the average crash rates at horizontal curved roads is approximately three times that of other road settings. To minimize the number of crashes occurring in horizontal curved roads, it is necessary to understand the risk factors that significantly affect the crashes.

In general, the risk factors of crashes while driving on horizontal curved roads consist of two categories: inside (e.g., in-vehicle secondary task engagements) and outside (e.g., road geometry or other vehicles) of the vehicle. Several studies have investigated the effects of in-vehicle secondary task engagements on driving performance on curved roads (e.g., Tsimhoni & Green, 2001; Tsimhoni, Smith, & Green, 2004; Lehtonen, Lappi, & Summala, 2012). One of the typical findings was that the secondary task significantly degraded driver's lane keeping performance on horizontal curved roads.

On the other hand, several studies have explored the relationship between driving performance and the risk factors from the outside of vehicle in horizontal curved road settings. For example, Godthelp (1986) and Ben-Bassat and Shinar (2011) found that the smaller the

curve's radii, the harder it is for a driver to maintain a stable lane position (i.e., higher standard deviations of lane position and steering wheel angle). Moreover, it was found that the higher the degree of curvature, the lower the mean speed of vehicle (Andjus & Maletin, 1998; Figueroa & Tarko, 2005). With respect to the effect of curve direction on driving performance, Bella (2013) found that speed and lateral position varied depending on the curvature (sharp vs. moderate) as well as direction (left vs. right). In addition, Calvi (2015) found that lateral positions were lower at left curve and higher at right curve.

However, relatively fewer studies have investigated the effect of lead vehicle on the safety of horizontal curve driving. Although many car-following driver models have been developed (e.g., Boer, Ward, Manser, Yamamura, & Kuge, 2005; Ossen & Hoogendoom, 2005; Yang & Peng, 2010) over the past decades, comparison between car-following and free-flow conditions has been seldom investigated.

In the study reported in this chapter, I investigated the effects of road geometry (i.e., road curvature and curve direction) and existence of lead vehicle on driving performance. In addition to their main effects, I investigated whether there are interaction effects between the factors on the driving performance.

6.2. Methods

6.2.1. Participants

Twenty-four participants (18 males and 6 females) between the ages of 19 and 33 (M = 23.6, SD = 4.1) were recruited through on-campus email advertisements to take part in this experiment. They were all licensed drivers who had average driving mileage of 32,960 miles. They were screened before their participation to ensure normal or corrected-to-normal vision.

6.2.2. Apparatus and driving scenarios

A fixed-based driving simulator was used to collect driving performance in horizontal curved roads. The simulator included a 24-inch LCD monitor (HP ZR24w) and a Logitech G27 RT racing wheel set. Using an open source driving simulation software, the Open Racing Car Simulator (TORCS), multiple driving scenarios including horizontal curved roads were created (see examples in Figure 6.1). The horizontal curved roads had four levels of curvature radius (100, 200, 400, and 800 m; all curve deflection angles were 180 degrees) in both right and left directions and they were connected by 300-meter transition straight roads to each other.

Rural highway roads with two lanes in one direction were used and had a flat surface with only horizontal curves and no vertical curves. The width of each lane was 5 meters. There were approximately 4-meter-wide shoulders on each side of the roads. The average driving duration for each scenario was approximately 10 minutes with an average speed of 60 km/h. A resolution of 1600 × 1200 pixels was used to generate the graphics of the driving scenarios. Both the driver and lead vehicles were a mid-size sedan (Peugeot 406). A dashboard including a speedometer and a tachometer was displayed at the center-bottom of the screen. A rear-view mirror was presented at the center-top of the screen. Some parts of these driving scenarios were used to evaluate a computational model that predicts eyes-off-road behavior and workload in performing different secondary tasks on horizontal curves (Jeong & Liu, 2017).

Driving performance data (e.g., speed, lane position, steering wheel angle, distance headway, and time headway) were automatically collected by the simulation system at the 60 Hz frequency.

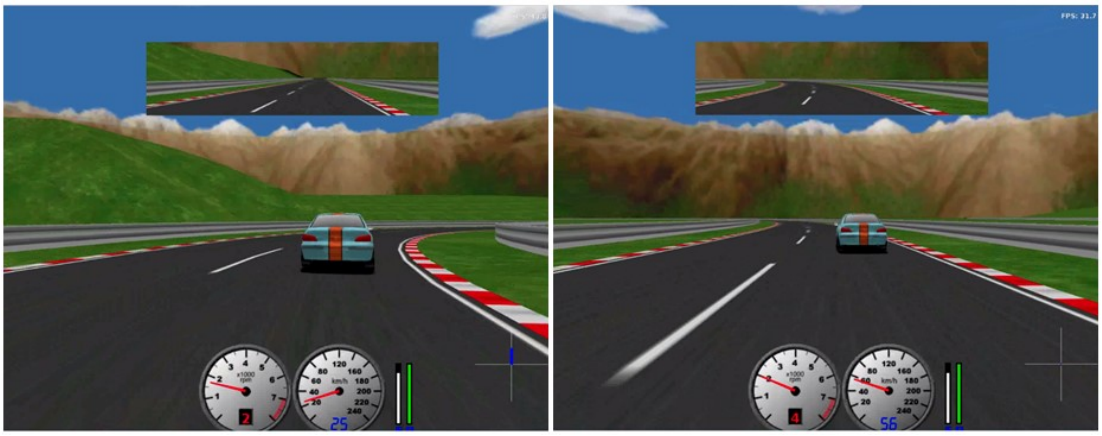


Figure 6.1 – Examples of driving scene for car-following task (sharp right curve on the left, moderate left curve on the right)

6.2.3. Experimental design

The experiment followed 2 × 2 × 2 mixed factor design with two levels of road curvature (sharp vs. moderate), two levels of curve direction (left vs. right), and two levels of driving condition (car-following vs. free-flow). All levels of each independent variable were randomly assigned to the participants. For road curvature, only two curvature radii were used: 100-meter curve as sharp, whereas 800-meter curve as moderate. A repeated measures analysis was performed for the following dependent variables: (1) standard deviations of lane position and steering wheel angle as measures of lane keeping performance, (2) vehicle speed as well as standard deviations of distance headway and time headway as measures of car-following performance.

6.2.4. Procedures and driving tasks

Upon arrival at the laboratory, participants were informed about the purposes of the experiment and asked to read and sign a consent form if they agreed to do so. They drove through a practice drive for 5 - 10 minutes to become familiar with the driving simulator. Participants performed two driving tasks: car-following and free-flow tasks. For the car-following task, participants were asked to follow a lead vehicle running with a constant speed

(i.e., 75 km/h) only on the right lane, while maintaining a safe car-following distance. For the free-flow task, participants were asked to drive with their own desired speed but obey the speed limit, 75 km/h. For both tasks, participants were instructed to drive on the right lane and maintain the vehicle in the center of the lane in a normal and safe manner. After each task, they were asked to complete a DALI (Driving Activity Load Index; Pauzié, 2008) survey and given a 5-minute break to avoid fatigue from each task. At the end of all the tasks, they completed a payment form.

6.2.5. Data analysis

The dependent measures of mean speed, standard deviations of lane position (SDLP) and steering wheel angle (SDSWA), standard deviations of distance headway and time headway were analyzed by repeated-measures three-way ANOVA (2 road curvatures × 2 curve directions × 2 driving conditions) using SPSS Statistics 24. The subjective driving workload collected by DALI was analyzed by one-way ANOVA for two driving conditions. Prior to performing the ANOVA tests, all the dependent variables were verified whether they violate the normality assumption of ANOVA, using a criteria of skewness (< 2) and kurtosis (< 7) (Curran, West, & Finch., 1996). Additionally, the Levene's test for homogeneity of variance was used to investigate the assumption of homogeneity of variance across groups. Effect sizes were measured by a partial eta-squared ($n_p{}^2$). The significant level for all statistical tests was $p \leq .05$.

6.3. Results

6.3.1. Mean speed

As shown in Figure 6.2, a repeated-measures ANOVA of mean speed revealed a significant main effect of driving condition (F (1, 136) = 407.4, $p < .001$, $n_p{}^2 = .75$) and road

curvature (F (1, 136) = 6.68, p = .011, $n_p{}^2$ = .05): higher in the Free-flow task (mean ($\bar{x}$) = 71.8, standard error of mean (*SEM*) = .98) than the Car-following task ($\bar{x}$ = 56.5, *SEM* = .24), higher at Moderate curve ($\bar{x}$ = 62.2, *SEM* = 1.06) than Sharp curve ($\bar{x}$ = 61.0, *SEM* = .93). The interaction between driving condition and road curvature was also significant (F (1, 136) = 8.64, p = .004, $n_p{}^2$ = .06).

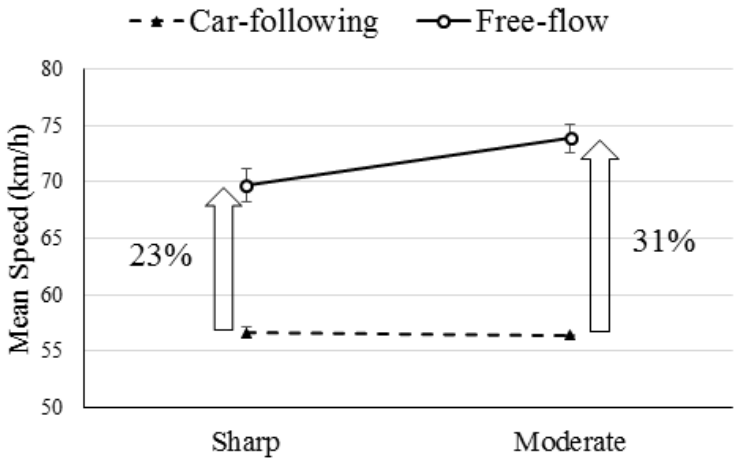


Figure 6.2 – Two-way interaction in Mean speed between the driving condition and road curvature (Error bars = *SEM*)

6.3.2. SDLP and SDSWA

As shown in Figure 6.3, an ANOVA for SDLP revealed a significant main effect of road curvature (F (1, 136) = 19.36, p < .001, $n_p{}^2$ = .13), but no significant main effects of other factors nor any interaction effect. The SDLP at Sharp curve ($\bar{x}$ = .32, *SEM* = .02) was higher than that at Moderate curve ($\bar{x}$ = .23, *SEM* = .01).

SDSWA was significantly affected by road curvature (F (1, 136) = 168.76, p < .001, $n_p{}^2$ = .55) and interaction between road curvature and curve direction (F (1, 136) = 3.95, p = .049 < .05, $n_p{}^2$ = .03; marginally significant). The SDSWA at Sharp curve ($\bar{x}$ = .37, *SEM* = .02) was higher than that at Moderate curve ($\bar{x}$ = .10, *SEM* = .005).

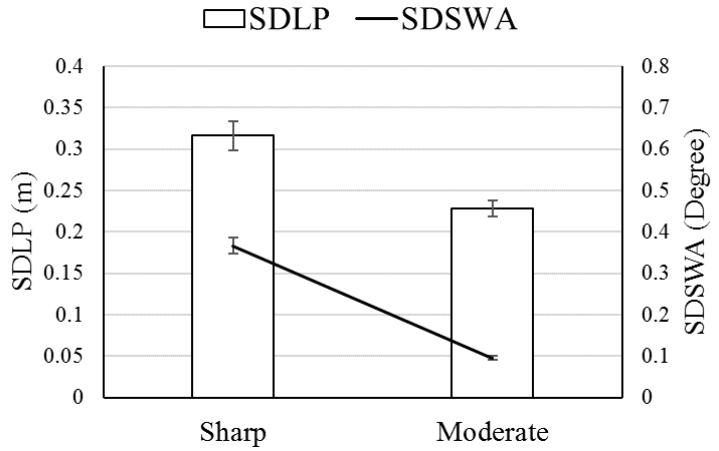


Figure 6.3 – Lateral control measures by road curvature
(Error bars = *SEM*)

6.3.3. Headway measures for the car-following task

The significant effect of road curvature (F (1, 92) = 6.50, p = .012, $n_p{}^2$ = .07) showed the higher SD distance headway at Moderate curve ($\bar{x}$ = 6.17, *SEM* = .66) than at Sharp curve ($\bar{x}$ = 3.93, *SEM* = .58), as shown in Figure 6.4. The SD time headway was also significantly affected by road curvature (F (1, 92) = 8.17, p = .005, n_p^2 = .08): higher at Moderate curve (x̄ = .41, *SEM* = .04) than Sharp curve ($\bar{x}$ = .25, *SEM* = .03).

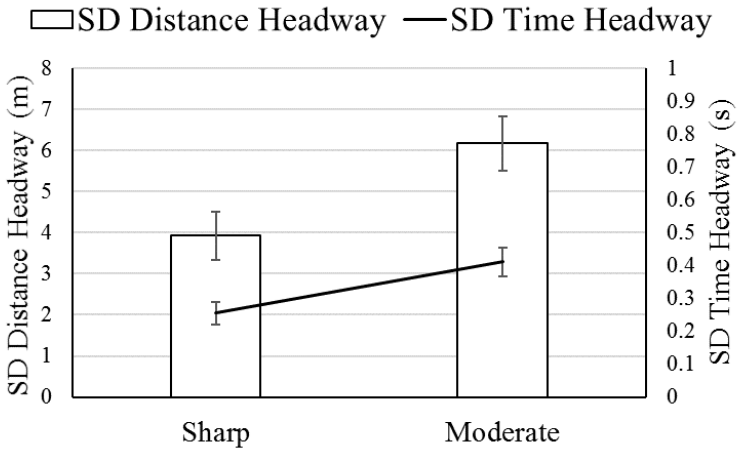


Figure 6.4 – The effect of road curvature on headway measures (car-following task)
(Error bars = *SEM*)

6.3.4. Subjective driving workload

Six driving activity load measures with a scale ranging from 1 (lowest) to 7 (highest) were analyzed: (1) effort of attention (i.e., level of attention required by the driving task), (2) visual demand (i.e., level of visual demand necessary for the driving task), (3) auditory demand (i.e., level of auditory demand necessary for the driving task), (4) temporal demand (i.e., level of specific constraint due to timing demand during the driving task), (5) interference (i.e., level of disturbance during the driving task), and (6) situational stress (i.e., level of constraint or stress during the driving task). Although there were differences of each driving workload measure between car-following and free-flow tasks (shown in Figure 6.5), no statistically significant differences were found (all $p > .05$).

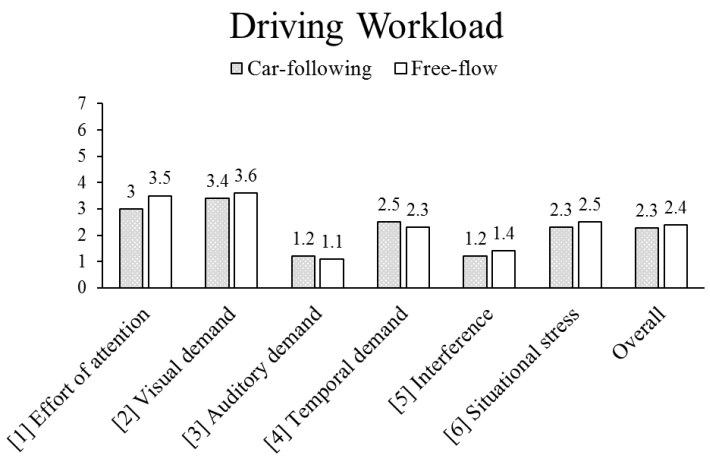


Figure 6.5 – Driving workload by driving condition

6.4. Discussion

6.4.1. Effects of road geometry

Significant effects of road curvature on driving performance were found: First, the SDLP was 39% higher at Sharp curve than at Moderate curve. Second, SDSWA was 3.7 times higher at

Sharp curve than Moderate curve. Lastly, the mean speed was 2% higher at Moderate curve than Sharp curve. These results are similar to the findings of previous studies (Godthelp, 1986; Andjus & Maletin, 1998; Figueroa & Tarko, 2005; Ben-Bassat & Shinar, 2011). In other words, the variability in lane keeping was higher and driving speed was reduced while driving on sharper curves.

For the car-following task, SD distance headway was 57% lower and SD time headway was 64% lower at Sharp curve than Moderate curve. The lower standard deviation of time and distance headway may indicate their driving strategy and tendency to keep a safe and constant following distance away in sharper curves.

In contrast to the previous studies (e.g., Bella, 2013; Calvi, 2015), the effect of curve direction on driving performance was not found in this study. This might be because the simulated driving roads used in this study had fewer roadside objects (e.g., vegetation, road signs, and guardrail barriers) than their studies, which might affect driver's perception of road and safety (Van Der Horst & De Ridder, 2007).

6.4.2. Effects of lead vehicle

The mean speed was 27% higher in free-flow than in car-following condition. Participants reduced their speed when a lead vehicle is ahead of them, whereas they increased their speed when they were free to drive. Moreover, the mean speed was significantly affected by the interaction between road curvature and driving conditions. In other words, the mean speed of two driving conditions varied depending on the road curvatures. The mean speed increased 23 % at Sharp curve and 31 % at Moderate curve, from car-following to free-flow condition. This result indicated that drivers increase their speed if there is no lead vehicle in front of them, even in the curved roads. The extent of speed increase is lower at the Sharp curve than Moderate

curve, probably in order to maintain the safe lane-keeping performance (nonetheless, the SDLP and SDSWA was higher at Sharp curve than Moderate curve).

Interestingly, however, there was no significant difference of subjective driving workload between two driving conditions. It should be noted that, while driver's self-evaluated workload showed no difference between the car-following and free-flow conditions, the driving performance (i.e., mean speed) varied by the driving conditions.

6.5. Conclusions

This study examined the effects of road geometry (i.e., curve direction and road curvature) and existence of lead vehicle on driving performance. The results showed that the degree of road curvature affected driving performance, especially for mean speed, variability of lane-keeping performance, and headway. Moreover, the mean speed was affected by both road curvature and whether the lead vehicle existed.

One limitation of this study is that the lead vehicle's speed was consistent and the movement was relatively static and steady-state. The major reason of using consistent speed for the lead vehicle was its simplicity to compare the driving performance between presence and absence of the lead vehicle. However, it would be useful to investigate the effects of lead vehicle's speed and behavior on horizontal curve driving performance and driving workload in future research. Another limitation is the simulated driving setting, which might provide participants unrealistic environment while performing tasks, compared to test-track or naturalistic settings. For example, the widths of lane and shoulder in the simulation are different than a standard roadway. Despite of the limitations, the findings from this study provides important empirical data that can be used for driving safety on horizontal curved roads.

References

Andjus, V., & Maletin, M. (1998). Speeds of cars on horizontal curves. *Transportation Research Record: Journal of the Transportation Research Board*, (1612), 42-47.

Bella, F. (2013). Driver perception of roadside configurations on two-lane rural roads: Effects on speed and lateral placement. *Accident Analysis & Prevention*, *50*, 251-262.

Ben-Bassat, T., & Shinar, D. (2011). Effect of shoulder width, guardrail and roadway geometry on driver perception and behavior. *Accident Analysis & Prevention*, *43*(6), 2142-2152.

Boer, E. R., Ward, N. J., Manser, M. P., Yamamura, T., & Kuge, N. (2005, June). Driver performance assessment with a car following model. In *Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (pp. 433-440).

Calvi, A. (2015). Does Roadside Vegetation Affect Driving Performance? Driving Simulator Study on the Effects of Trees on Drivers' Speed and Lateral Position. *Transportation Research Record: Journal of the Transportation Research Board*, (2518), 1-8.

Curran, P. J., West, S. G., & Finch, J. F., (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis, *Psychological Methods, 1*(1), 16-29.

Godthelp, H. (1986). Vehicle control during curve driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *28*(2), 211-221.

Federal Highway Administration, Low Cost Treatments for Horizontal Curve Safety, FHWASA-07-002 (Washington, D.C.: 2006). Available at: http://safety.fhwa.dot.gov/roadway_dept/horicurves/fhwasa07002.

Figueroa, A., & Tarko, A. (2005). Speed factors on two-lane rural highways in free-flow conditions. *Transportation Research Record: Journal of the Transportation Research Board*, (1912), 39-46.

Jeong H., & Liu, Y. (2017). Modeling of Stimulus-Response Secondary Tasks with Different Modalities while Driving in a Computational Cognitive Architecture. In *Proceedings of the Ninth International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (pp.58-64).

Lehtonen, E., Lappi, O., & Summala, H. (2012). Anticipatory eye movements when approaching a curve on a rural road depend on working memory load. *Transportation Research Part F: Traffic Psychology and Behaviour*, *15*(3), 369-377.

Ossen, S., & Hoogendoorn, S. (2005). Car-following behavior analysis from microscopic trajectory data. *Transportation Research Record: Journal of the Transportation Research Board*, (1934), 13-21.

Pauzié, A. (2008). A method to assess the driver mental workload: The driving activity load index (DALI). *IET Intelligent Transport Systems*, *2*(4), 315-322.

Tsimhoni, O., & Green, P. (2001, October). Visual demand of driving and the execution of display-intensive in-vehicle tasks. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 45, No. 23, pp. 1586-1590). Sage CA: Los Angeles, CA: SAGE Publications.

Tsimhoni, O., Smith, D., & Green, P. (2004). Address entry while driving: Speech recognition versus a touch-screen keyboard. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(4), 600-610.

Torbic, D. J., Harwood, D. W., Gilmore, D. K., Pfefer, R., Neuman, T. R., Slack, K. L., & Hardy, K. K. (2004). A guide for reducing collisions on horizontal curves. *NCHRP Report*, *500*(7).

Van Der Horst, R., & De Ridder, S. (2007). Influence of roadside infrastructure on driving behavior: driving simulator study. *Transportation Research Record: Journal of the Transportation Research Board*, (2018), 36-44.

Yang, H. H., & Peng, H. (2010). Development of an errorable car-following driver model. *Vehicle System Dynamics*, *48*(6), 751-773.

Chapter 7.

Queuing Network Modeling of Driver Lateral Control on Curved Roads with Integration of Vehicle Dynamics and Reference Trajectory Tracking

7.1. Introduction

Nearly 25 percent of drivers who die each year on the American roadways are due to fatal crashes on curved roads (McGee & Hanscom, 2006). Therefore, it is important to study and understand driver lateral (or steering) control on curved roads in order to help reduce or avoid the crashes. In this regard, computational modeling of driver lateral control on curves can play a significant role in supporting quantitative analysis of driver's performance.

Since the last few decades, a variety of modeling studies have been conducted to quantify driver behavior in lateral control. One method used was based on control theory that assumes the human driver is one of the control elements in the driver-vehicle system (e.g., MacAdam, 2003). Another method used for the lateral control modeling was the driver preview model based on imitating drivers' preview/predictive behaviors (e.g., Ungoren & Peng, 2005). Researchers have also started to model driver performance using task-independent cognitive architectures, based on experimental psychology and neuroscience findings. Examples of cognitive architecture based driver model include the Adaptive Control of Thoughts-Rational (ACT-R) (e.g., Salvucci, 2006), and the QN-MHP (e.g., Liu, Feyen, & Tsimhoni, 2006).

In this study, I used the QN-MHP cognitive architecture to model driver lateral control on curved roads. The QN-MHP is a simulation model of cognitive processing system based on the

queueing network theory of human performance (see Liu et al., 2006 for more details). The QN-MHP architecture is composed of three subnetworks (perceptual, cognitive, and motor) and each subnetwork consists of multiple servers representing the functional components of the brain and body for human performance. The servers are connected by routes, while entities travel through the routes. One of the merits of using the QN-MHP is that it allows more than one server to act either in parallel or in serial. Thus, it is possible to model human performance in multi-task scenarios represented as multiple flows of entities, such as the driving performance under multi-task conditions.

In this study, a novel method, called reference trajectory tracking, was used to control vehicle's lateral motion in order to obtain higher accuracy of modeling both the position and time elements. This method was originally used to design autonomous vehicle's lateral movement by minimizing the spatial and temporal errors from the reference trajectory (e.g., Aguiar & Hespanha, 2007; Talj, Tagne, & Charara, 2013). Using the reference trajectory tracking concept (i.e., by making the virtual vehicle follow the built-in reference trajectory), I have developed a model to simulate a vehicle's lateral movement on curved roads with multiple levels of radius of road curvature. After model development for the driver's lateral control, model validation was conducted with the existing experimental data from Tsimhoni & Green (2003), with lateral control measurements such as steering wheel angle.

7.2. Computational modeling of lateral control

The model built in this study combines the original QN-MHP architecture with the road curvature information and vehicle dynamics for modeling driver's lateral control in driving on curved roads (See Figure 7.1). The lateral control model is implemented in MATLAB-Simulink

and has four main components: (1) road information, (2) vehicle dynamics, (3) visual perception, and (4) cognition & motor controls. As shown in Figure 7.1, entities carrying road information enter the QN-MHP architecture as well as the vehicle dynamics component:

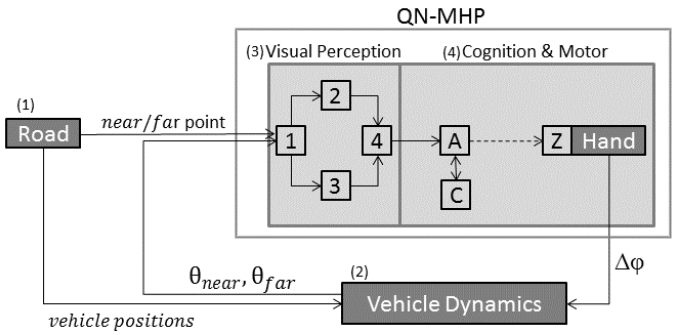


Figure 7.1 – Structure of four components with the QN-MHP

At the QN-MHP architecture, first, the entities enter the visual perception subnetwork: Servers 1 (visual input) → 2/3 (visual recognition/localization) → 4 (visual integration). The entities collected in Server 4 move to the cognitive & motor subnetworks by entering through Server A (visuospatial sketchpad) and exiting through Server Z (Actuators; connected to the hand server in this study). Once the entities arrive at the hand server, a signal is sent to the vehicle dynamics component so that the component can prepare the settings for carrying out the lateral control (e.g., yaw angle adjustment and near/far angle acquisition). At Server C (central executor), driver's steering wheel angle (φ) is determined (by Equation 5), based on the near/far angle obtained from the vehicle dynamics component. In this study, Server C was used (rather than the combination with Server F) to perform the driving lateral control because it was assumed that adjusting the steer wheel on curved road having a constant radius of curvature is a simple cognitive process, compared to such complex cognitive activities, mostly performed through Server F (e.g., multiple-choice comparison and decision, and math calculation).

On the other hand, the entities that travel from the road information component to the vehicle dynamics component carry road information (e.g., radius of road curvature and 2-dimentional coordinates of vehicle's reference trajectory), which are used to predict the yaw angle and vehicle's position, then eventually near angle and far angle (by Equations 3 and 4) at the vehicle dynamics component. The output (φ) determined in Server C is used for driver's lateral control when the entities arrive at the vehicle dynamic component in the next cycle after circulating through Server A and Server Z (and the hand server).

7.2.1. Road information component

One of the major factors that affects driver's visual perception on the curved road is the radius of road curvature (Dickmanns & Zapp, 1987; Shinar, Rockwell, & Malechi, 1980). In geometry, the radius of curvature, *R*, at a particular point is defined as the radius of the most approximate circle touching the point. With the assumption that the curve is differentiable up to the second order, the formula for the radius of curvature at any point *x* for the curve $z = f(x)$ is given by:

$$R = \left| \frac{(1+(\frac{dz}{dx})^2)^{3/2}}{\frac{d^2z}{dx^2}} \right| \quad (1)$$

where *x* is the lateral coordinate and *z* is the longitudinal coordinate (Do Carmo, 1976). Although this radius varies as the vehicle moves along the curve (Hastie & Stuetzle, 1989), it was assumed that the radius of road curvature in the current model indicates the radius when the vehicle is at the middle of curve. The road curvature parameter (*R*) is used as an input to estimate the yaw angle (ψ) at the vehicle dynamics component. The built-in reference trajectory data are used to determine the 2-dimensional geometrical center of vehicle position at the vehicle dynamics component. The details are described in the vehicle dynamics component section.

7.2.2. Vehicle Dynamics Component

In this component, three vehicle dynamics measurements are determined: vehicle positions, yaw angles, and near/far angles. The vehicle positions (i.e., lateral and longitudinal positions) are determined by minimizing the time/space-based errors from the built-in reference trajectory data, using interpolation method. Since the reference trajectory data are designated every 100 msec and the current simulation cycle time is every 50 msec, it is necessary to find the closest cycle time of reference trajectory; the trajectory data at the closest cycle time are regarded as the current vehicle positions.

In parallel, yaw angles are determined by the simulation cycle time. Yaw motion is one of the significant elements for controlling lateral movement in vehicle dynamics (Ackermann & Bünte, 1997; Rajamani, 2012). Yaw angle (ψ) is defined as the angle between the direction of the vehicle heading and the direction of the lane center. In the current study, the yaw angle was determined with the finding from Rajamani (2012):

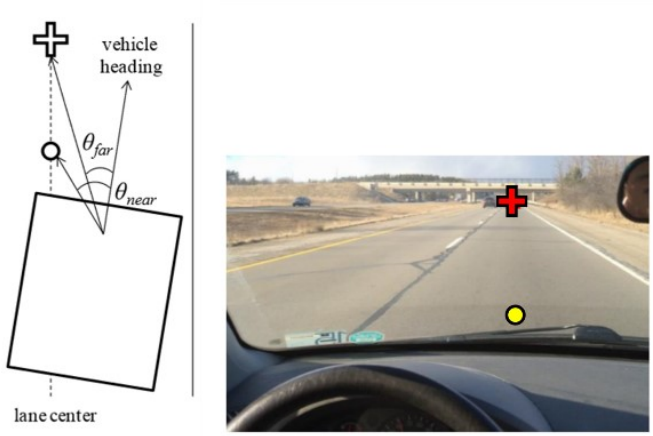
$$\psi_t - \psi_{t-1} = \frac{v_t \cdot \Delta t}{R} \quad (2)$$

v_t is the vehicle speed at time *t*, whereas *R* is radius of road curvature. Δt is the time elapsed from last cycle. Each cycle time depends on two time components: inter-arrival time and server processing time. The inter-arrival time is a fixed time, currently set as 50 msec as the default value for the visual stimulus generation rate. The server processing time is set as a shifted exponential distribution with mean *a* and an axis shift *b*. It is written as *E(a) + b*: the values of parameter <*a, b*> in this study were set as <17, 25> for perceptual servers, <12, 6> for cognitive servers, and <14, 10> for motor servers, based upon Feyen (2002). The present model steers the vehicle at a fixed speed of 72 km/h (or 45 mph). Using the lateral vehicle position (X_t) and yaw

angle (ψ_t) at each cycle time, near/far angles (i.e., the results of Equations 3 and 4) are determined at this vehicle dynamics component.

7.2.3. Visual perception component

In the current model, the near point and far point were generated every 50 msec and they were used as visual stimuli inputs at the visual perceptual servers (Servers 1 - 4). The near point indicates a visible point in front of the vehicle that the driver uses for estimating how adjacent the vehicle is to the center of lane, whereas the far point represents a visible point in front of vehicle that the driver uses to estimate a near future position (See Figure 7.2).



(a) Bird's-eye view (b) Driver's point-of-view

Figure 7.2 – (a) Near/far angles and (b) near/far points (a yellow dot and a red cross, respectively) (Feng, 2015)

Figure 7.2 illustrates the near angle and far angle: the near angle refers to the direction from the vehicle pointing to the near point relative to the direction of the vehicle heading, while the far angle represents the direction from the vehicle pointing to the far point relative to the direction of the vehicle heading.

The near angle and the far angle are determined (by using Equations 3 and 4 below) at the vehicle dynamics component after the input data for the component are obtained, such as the lateral coordinates of vehicle position (X_t) and yaw angles (ψ_t). While *nearDistance* indicates

the near point location (in distance) on the road (set as a constant, 10 m), *farTHW* represents the far point location (in time headway) on the road (set as a constant, 4 s).

$$\theta_{near,t} = \tan^{-1}\left(\frac{X_t}{nearDistance}\right) + \psi_t \quad (3)$$

$$\theta_{far,t} = \tan^{-1}\left(\frac{X_t}{farTHW \cdot v_t}\right) + \psi_t \quad (4)$$

7.2.4. Cognition and motor controls component

Once the near and far angles are stored in the cognitive subnetwork, steering wheel angle adjustment is conducted at Server C. The present model uses a steering wheel angle (φ) equation formulated by Salvucci (2006).

$$\Delta\varphi = k_{far} \cdot \Delta\theta_{far} + k_{near} \cdot \Delta\theta_{near} + k_l \cdot \min(\theta_{near}, \theta_{near_{max}}) \cdot \Delta t \quad (5)$$

In which: $\theta_{near_{max}}$ (set as a constant, 0.07 radian) is for limiting the contribution of the θ_{near} to changes in steering wheel angle. k_{far}, k_{near}, and k_l indicate the weights for the three terms (set as 7, 4, and 3, after the multiple validations to obtain similar results with the experimental data). The motor server (indicated as Z in Figure 7.1) sends a signal to the hand server so turning actions are performed in the vehicle dynamic component using the data of the change of steering wheel angle (Δφ) in each cycle. The turning action time (i.e., steering time), is taken, based on the estimation with a steering rate of 963 degree/sec from Forkenbrock & Elsasser (2005).

7.3. Validation results

The lateral control driving model was evaluated using the empirical data of Tsimhoni & Green (2003), conducting a driving simulation experiment with 24 participants (12 younger (M = 23, SD = unknown); 12 older (M = 68, SD = unknown)). They drove on a 3.6 m wide single-lane road made up of a curve (two levels: R = 200 and 400 m) connecting two straight roads before

and after of the curved road. The participants were asked to drive with a constant cruise controlled speed (72 km/h or 45 mph).

Figure 7.3-(a) shows the mean yaw angles generated by the vehicle dynamics module (in degree; during 10 runs of simulation) during driving on four different radii of road curvature. The mean yaw angles are fairly constant over time, and the value decreases as the vehicle is driven on a curve with a larger radius. Similarly, the mean steering wheel angle (in degree; during 50-second driving) decreases as the driver model drove on larger-radius curves (See Figure 7.3-(b)).

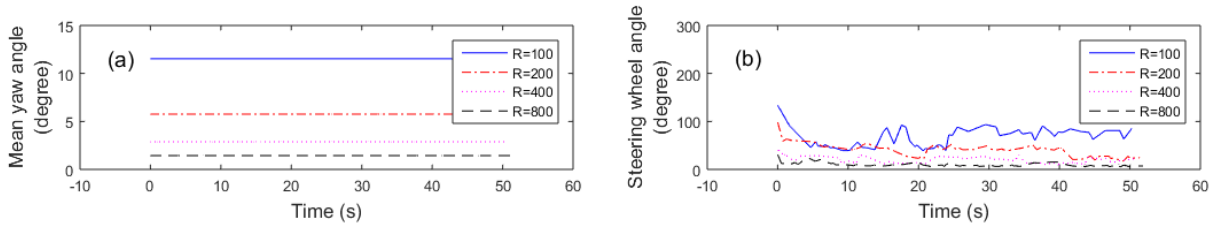


Figure 7.3 – (a) Mean yaw angle and (b) steering wheel angle of simulation results (in degree) in four different radii of road curvature at 72 km/h

With regard to the comparison between simulation and experimental results, as shown in Figure 7.4, the mean steering wheel angles of the driver model quite closely fit the experimental data for both radii of road curvature (R = 200 and 400 m).

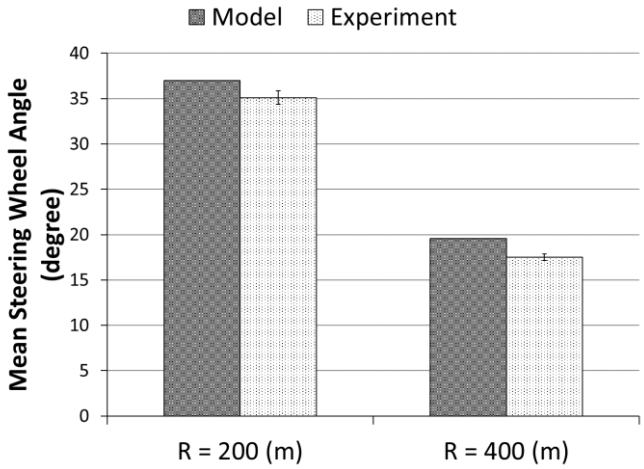


Figure 7.4 – Mean steering wheel angle comparisons between the QN simulation and experimental results

7.4. Discussion

Using the vehicle dynamics and the method of reference trajectory tracking, as seen in Figure 7.4, the model simulation was able to generate mean steering wheel angle data that were fairly similar to the human experimental study. Furthermore, since the vehicle in simulation was to be followed the built-in reference trajectory, the lateral gap between the center of vehicle and road was close to zero.

In Tsimhoni & Liu (2003), a driver steering model was also successfully developed using processing logics including detecting orientation, selecting a steering strategy, and steering action: the model yielded the steering angle and lateral position in two fixed radii of curvature, similar to the experimental data using human subjects. However, one major contribution of this current study is that the method of reference trajectory tracking can help prediction of the driving performance on different curved conditions (rather than just fixed radius of curvature). In other words, once having any built-in reference trajectory data (including vehicle coordinates at each cycle time), it will be possible to model the vehicle control on that trajectory.

Some limitations of the current study are that: (1) the current lateral vehicle control model ran with a fixed speed; (2) only the mean steering wheel angle measurement has been validated due to the lack of corresponding human experimental data for validation. I am extending the present model so that it can run in more complicated settings, including a longitudinal speed control, with or without a lead vehicle, and so forth. Moreover, using the reference trajectory tracking method which can demonstrate vehicle's movement in three-dimensional space (Aguiar & Hespanha, 2007), I plan to model driver elevation control on the upward/downward slopes.

References

Ackermann, J., & Bünte, T. (1997). Yaw disturbance attenuation by robust decoupling of car steering. *Control Engineering Practice*, *5*(8), 1131-1136.

Aguiar, A.P. and Hespanha, J. P. (2007). Trajectory-tracking and path-following of underactuated autonomous vehicles with parametric modeling uncertainty. *IEEE Transactions on Automatic Control*, *52*(8), 1362-1379.

Dickmanns, E. D., & Zapp, A. (1987, February). A curvature-based scheme for improving road vehicle guidance by computer vision. *In Cambridge Symposium_Intelligent Robotics Systems* (pp. 161-168). International Society for Optics and Photonics.

Do Carmo, M. P. (1976). *Differential geometry of curves and surfaces* (Vol. 2). Englewood Cliffs: Prentice-hall.

Feng, R. (2015). *Queuing Network Modeling of Human Multitask Performance and its Application to Usability Testing of In-Vehicle Infotainment Systems* (Doctoral dissertation, University of Michigan).

Feyen, R. G. (2002). Modeling human performance using the queuing network-model human processor (QN-MHP) (Doctoral dissertation, University of Michigan).

Forkenbrock, G. J., & Elsasser, D. (2005). An assessment of human driver steering capability. *National Highway Traffic Safety Administration DOT HS* 809, 875.

Hastie, T., & Stuetzle, W. 1989. Principal curves. *Journal of the American Statistical Association*, *84*(406), 502-516.

Jeong, H., & Liu, Y. (2016, September). Computational Modeling of Finger Swipe Gestures on Touchscreen Application of Fitts' Law in 3D Space. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 60, No. 1, pp. 1721-1725). SAGE Publications.

Lim, J. H., Liu, Y., & Tsimhoni, O. 2010. Investigation of driver performance with night-vision and pedestrian-detection systems—Part 2: Queuing network human performance modeling. *IEEE Transactions on Intelligent Transportation Systems*, *11*(4), 765-772.

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.

MacAdam, C. C. (2003). Understanding and modeling the human driver. *Vehicle System Dynamics*, *40*(1-3), 101-134.

McGee, H. W., & Hanscom, F. R. (2006). *Low-cost treatments for horizontal curve safety* (No. FHWA-SA-07-002).

Rajamani, R. (2012). Lateral vehicle dynamics. *In Vehicle Dynamics and control* (pp. 15-46). Springer US.

Salvucci, D. D. (2006). Modeling driver behavior in a cognitive architecture. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *48*(2), 362-380.

Shinar, D., Rockwell, T. H., & Malechi, J. A. (1980). The effects of changes in driver perception on rural curve negotiation∗. *Ergonomics, 23(*3), 263-275.

Talj, R., Tagne, G., & Charara, A. (2013, July). Immersion and invariance control for lateral dynamics of autonomous vehicles, with experimental validation. In *European Control Conference (ECC'13)* (pp. 968-973).

Tsimhoni, O., & Green, P. (2003). Time-sharing of a visual in-vehicle task while driving: The effects of four key constructs. In *Proceedings of the Second International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design. Park City, Utah. July* (pp. 113-118).

Tsimhoni, O., & Liu, Y. (2003, October). Modeling Steering Using the Queueing Network—Model Human Processor (QN-MHP). In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 47, No. 16, pp. 1875-1879). SAGE Publications.

Ungoren, A. Y., & Peng, H. (2005). An adaptive lateral preview driver model. *Vehicle system dynamics, 43*(4), 245-259.

Wu, C., & Liu, Y. 2008. Queuing network modeling of transcription typing. *ACM Transactions on Computer-Human Interaction (TOCHI), 15*(1), 6.

Chapter 8.

Development and Evaluation of a Computational Cognitive Model for In-vehicle Direct/Indirect Manual and Speech Interactions

8.1. Introduction

To enhance drivers' safety and convenience in interacting with in-vehicle technologies, a variety of in-vehicle information systems (IVISs) using direct and indirect input devices have been developed (Harvey, Stanton, Pickering, McDonald, & Zheng, 2011a; Green, 2000; Stevens, Quimby, Board, Kersloot, & Burns, 2002). Direct input devices (e.g., touchscreens, light pens, and voice input) do not require any translations between actions of the operator and the device, whereas indirect input devices (e.g., joysticks, trackballs, and rotary knobs) require the translations between actions of the operator and the device (Jacob, 1996; Rogers, Fisk, McLaughlin, & Pak, 2005; McLaughlin, Rogers, & Fisk, 2009). Examples of in-vehicle direct devices include touchscreens on the center console (e.g., *MyFord Touch)* and speech interactions systems (e.g., *Ford Sync*). Examples of in-vehicle indirect devices are steering wheel-mounted controls (e.g., cruise control buttons in commercial vehicles) and remote rotary knob controls (e.g., *Lexus Remote Touch* or *BMW iDrive Controller*).

Many previous automotive studies have found that driver's visual distraction (i.e., visual attention away from the roadway) is one of the critical factors that increase unsafe driving performance and car crash risk (Ranney, Mazzae, Garrott, & Goodman, 2000; Klauer, Dingus, Neale, Sudweeks, & Ramsey, 2006). In particular, when drivers are involved in secondary tasks

using visual-manual interfaces, shorter and less frequent glances are focused on the road, compared to using speech (or voice) interfaces (Shutko, Mayer, Laansoo, & Tijerina, 2009; Tsimhoni, Smith, & Green, 2004). To minimize visual distraction caused by the visual-manual tasks, most of the motor vehicle manufacturers have developed speech interfaces and integrated them into their in-vehicle systems. For example, Ford SYNC with MyFord Touch is one of the best-known in-vehicle communication systems, including speech interfaces, cluster displays with associated steering wheel control, and touchscreen (Shutko & Tijerina, 2011; Lo & Green, 2013). The speech interfaces can allow drivers to call a person in a phone book list, request a song, and operate navigation and climate control systems, and so on.

Along with developing these IVISs using advanced automotive technologies, it is also necessary to evaluate the usability of the IVISs before they are commercialized. Traditional methods for evaluating the usability of the products include end-user tests in the laboratory setting, heuristic evaluation, and survey (Nielsen, 1993; Dumas & Redish, 1999; Shneiderman, 2010; Harvey, Stanton, Pickering, McDonald, & Zheng, 2011b; Ocak & Cagiltay, 2017). However, these methods typically involve creating physical prototypes, recruiting human subjects or evaluation experts, as well as coordinating the usability experiments; thus, they require costly and time-consuming efforts to conduct. On the other hand, cognitive modeling methods generally use the simulation of human behavior and performance based on scientific and engineering theories and knowledge; therefore, they do not require physical prototypes, usability experience expertise, and human subjects for the usability tests.

In this chapter, I introduce a computational cognitive model study for in-vehicle direct/indirect manual and speech interaction systems' usability test. The model was developed using the Queueing Network (QN) cognitive architecture (Liu, Feyen, & Tsimhoni, 2006) and

evaluated with human subject data from Chen, Tonshal, Rankin, & Feng (2016).

8.2. Background

Table 8.1 summarizes existing literature of cognitive modeling studies to predict driver behavior and performance while using in-vehicle interface systems. All the studies reported in Table 8.1 were involved in modeling the use of in-vehicle direct-manual devices (e.g., touchscreens or physical keypads). However, only a few of them (Schneegaß, Pfleging, Kern, & Schmidt, 2011; Feng, Liu, & Chen, 2017; Purucker, Naujoks, Prill, & Neukum, 2017) dealt with indirect-manual controls (i.e., use of rotary knobs) or voice controls (Salvucci, 2009). Almost all of them modeled human performance when the vehicle was moving (namely, in the driving mode).

However, there have been no studies that predict human behavior for both the direct/indirect manual and voice input controls in a single model. Moreover, none of the previous studies modeled the use of steering wheel-mounted buttons, a widely used indirect-manual (or remote-manual) device.

Table 8.1 – Cognitive modeling studies for in-vehicle interfaces

Studies	Based cognitive architecture	Measures validated	Task type: Manual (Direct/Indirect manual) / Voice	Driving mode	Validation data source and sample size (N)
This chapter	QN-MHP	Task completion time, Mental workload	Yes (touchscreen/5-way) / Yes	Yes	Chen et al. (2016) (N = 35)
John & Salvucci (2005)	KLM + ACT-R	Task completion time, Lateral deviation	Yes (touchscreen/no) / No	Yes	Own (N = unknown)
Pettitt, Burnett, & Stevens (2007)	KLM	Total Shutter Open Time (TSOT), static task time	Yes (touchscreen/no) / No	Yes	Own (N = 12)
Salvucci (2009)	ACT-R (Distract-R)	Task time, Lateral velocity	Yes (physical keypad/no) / Yes	Yes	Salvucci (2001) (N = 16)
		Lateral velocity	Yes (physical keypad/no) / No	Yes	Reed & Green (1999) (N = 12)
		Task time, heading error, Detection error	Yes (physical keypad/no) / Yes	Yes	Greenberg et al. (2003) (N = 48)
Schneegaß et al. (2011)	KLM	Task completion time	Yes (physical keypad/rotary knob) / No	Yes	Own (N = 10)
Harvey & Stanton (2013)	Critical path analysis	Task completion time	Yes (touchscreen/no) / No	No	Own (N = 20)
Feng et al. (2017)	QN-MHP	Task completion time, Total eyes-off-road time, Mental workload	Yes (touchscreen, physical keypad/rotary knob) / No	Yes	Own (N = 20)
Purucker et al. (2017)	KLM	Total eyes-off-road time (TEORT)	Yes (touchscreen/rotary knob) / No	Yes	Own (N = 18)

Using the QN-MHP architecture, Feng et al. (2017) recently developed a cognitive model to predict human performance for direct/indirect manual input tasks using an IVIS. In this chapter, I describe the research work that extended the functionality of their QN-model by additionally enabling it to predict human performance for (1) voice input tasks and (2) advanced indirect-manual input tasks (i.e., steering wheel-mounted control).

8.3. Model Development

8.3.1. Task Analysis

An in-vehicle task, called "*Deflate lower lumbar from 50 to 40*" studied by Chen et al. (2016) was used in this study. This task is to decrease the comfort level from 50 to 40 (note that each click decreases 10 levels), using three different methods (i.e., two manual control methods and a voice control method). Table 8.2 shows the brief procedures to complete the task, depending on the methods used. Direct Manual Control is an input method to click target buttons on the touchscreen display located on the center stack in the car. Indirect Manual Control is to press the physical buttons on the steering wheel-mounted control while confirming visual feedback on the cluster display located on the dashboard. Voice Control is to use a voice command, "deflate lower lumbar".

Table 8.2 – The task procedures of "*Deflate lower lumbar from 50 to 40*"

Method	Direct-Manual Control (touchscreen buttons)	Indirect-Manual Control (steering-mounted buttons)	Voice Control
Step 1	Click 'Settings'	Press '↓' button × 3 times	Say "deflate lower lumbar"
Step 2	Click 'Vehicle'	Press 'OK' button	–
Step 3	Click 'Seat Comfort'	Press '↓' button × 2 times	–
Step 4	Click '-' (deflation) on lower lumbar	Press '←' button	–

To perform the QN-MHP modeling, task analysis was conducted using the NGOMSL (Natural Goals, Operators, Methods, and Selection rules Language)-style task description (Kieras, 1999). Three sample results of the task analysis for "Direct-Manual", "Indirect-Manual", and "Voice" input subtasks are shown in Table 8.3.

Table 8.3 – NGOMSL-style description of typical “Direct-Manual”, “Indirect-Manual”, and “Voice” input subtasks

Goal: Click a button on touchscreen	Goal: Press a 5-way button and confirm its visual feedback on dashboard-cluster	Goal: Say a voice command
*TC 1: *Look-at* <target type> on touchscreen at location <x, y>	TC 1: *Look-at* <target type> on 5-way buttons at location <x, y>	TC 1: *Look-at* <target type> on touchscreen at location <x, y>
TC 2: *Store* the <target value> on †STM	TC 2: *Store* the <target value> on STM	TC 2: *Store* the <target value> on STM
TC 3: *Retrieve* the <target value> from STM	TC 3: *Retrieve* the <target value> from STM	TC 3: *Retrieve* the <target value> from STM
TC 4: *Compare* <target value> to <expected value>	TC 4: *Determine-thumb-movement(-for-reaching)*	TC 4: *Compare* <target value> to <expected value>
TC 5: *Decide* If results = 1, go to TC 6, else return results = 0	TC 5: *Reach-with-thumb*	TC 5: *Decide* If results = 1, go to TC 6, else return results = 0
TC 6: *Look-at* <target type> on touchscreen at location <x, y>	TC 6: *Look-at* <target type> on 5-way buttons at location <x, y>	TC 6: *Say* <syllables>
TC 7: *Store* the <target value> on STM	TC 7: *Store* the <target value> on STM	TC 7: Return with goal accomplished
TC 8: *Retrieve* the <target value> from STM	TC 8: *Retrieve* the <target value> from STM	
TC 9: *Determine-hand-movement*	TC 9: *Determine-thumb-movement(-for-clicking)*	
TC 10: *Reach-with-hand*	TC 10: *Click-with-thumb*	
TC 11: *Look-at* <target type> on touchscreen at location <x, y>	TC 11: *Look-at* <target type> on dashboard-cluster at location <x, y>	
TC 12: *Store* the <target value> on STM	TC 12: *Store* the <target value> on STM	
TC 13: *Retrieve* the <target value> from STM	TC 13: *Retrieve* the <target value> from STM	
TC 14: *Determine-finger-movement*	TC 14: *Compare* <target value> to <expected value>	
TC 15: *Click-with-finger*	TC 15: *Decide* If results = 1, go to TC 11, else return results = 0	
TC 16: Return with goal accomplished	TC 16: Return with goal accomplished	*TC = Task component, †STM = Short-term memory

8.3.2. Operator Developments

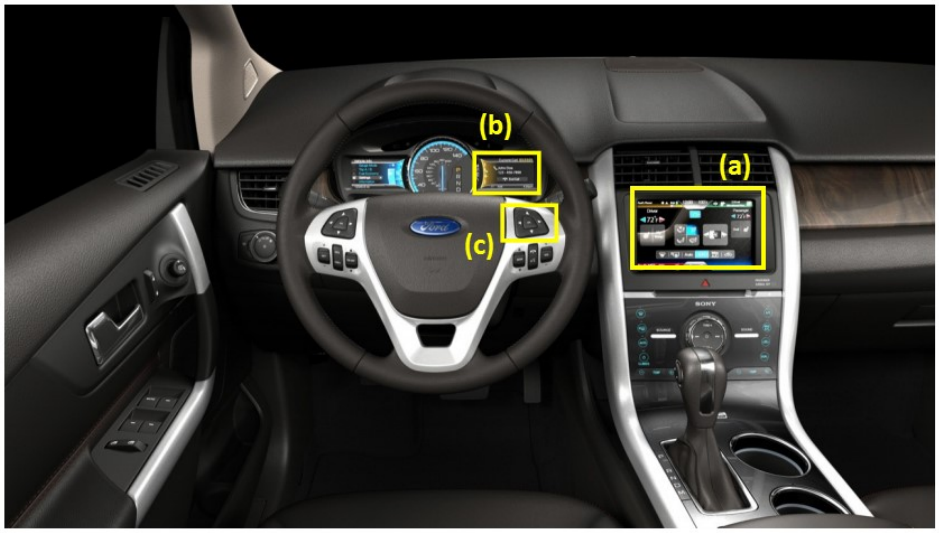
In the recent QN-MHP development, in addition to visual-manual-related operators (e.g., Feng, Liu, Chen, Filev, & To, 2014; Jeong & Liu, 2016, 2017c), several voice-related operators (e.g., *Say(a-syllable)*, *Listen-to(a-syllable)*, and *Listen-to(a-tone)*) have been developed to implement any task using auditory stimuli and speech responses (Jeong & Liu, 2017a, 2017b). In the current study, I adopted the major operators from these previous studies and extended the functions of the *Say(a-syllable)* operator so that the model can implement a variety of voice commands used in the practical driving settings. To do so, a pre-determined voice library module (example lists are shown in Table 8.4) was used. This module functions as an intermediary between human voice commands and the QN-MHP model, and it was created based upon the button information on the device. The voice command execution time (i.e., 130 – 170 msec per syllable, depending on the practiced level) was determined by the related finding of John (1990). The level of 130 msec per syllable was used, under the assumption that speaking the voice commands is highly practiced. For example, the voice command execution time for "*de-flate-low-er-lum-bar*" is 6 × 130 = 780 msec.

Table 8.4 – Example lists in the voice library module

Pronounced words by syllable	Button information (e.g., words, numbers, and symbols)
phone	Phone
help	?
nav-i-ga-tion	Navigation
en-ter-tain-ment	Entertainment
ei-em	AM
one	1
ze-ro	0
in-flate-low-er-lum-bar	+
de-flate-low-er-lum-bar	−
in-flate-up-per-lum-bar	+
de-flate-up-per-lum-bar	−

8.3.3. Digital Mockup Development

Using MATLAB GUIDE (Graphical User Interface Design Environment), three digital mockups of IVIS devices (shown in Figure 8.1) were developed and used as a prototyping tool: (a) center stack touchscreen device − for direct-manual and voice controls; (b) dashboard-cluster − for indirect-manual control; and (c) steering-wheel-mounted 5-way buttons − for indirect-manual control.

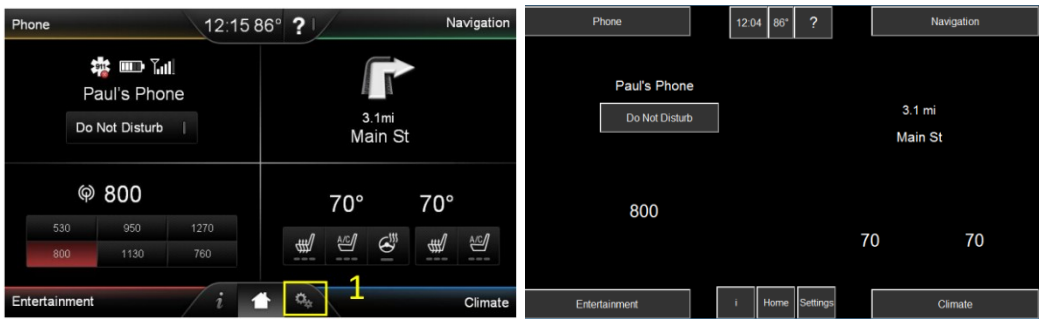


(a) center stack touchscreen, (b) dashboard-cluster, (c) steering wheel-mounted 5-way buttons

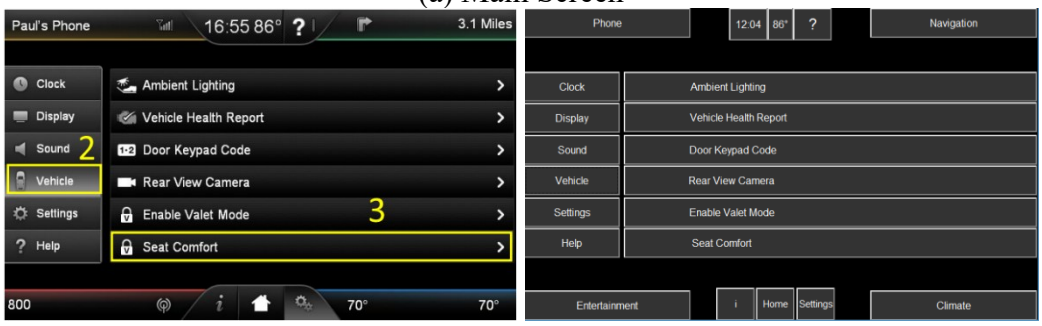
Figure 8.1 – Three typical devices for IVISs

8.3.3.1. Center stack touchscreen device

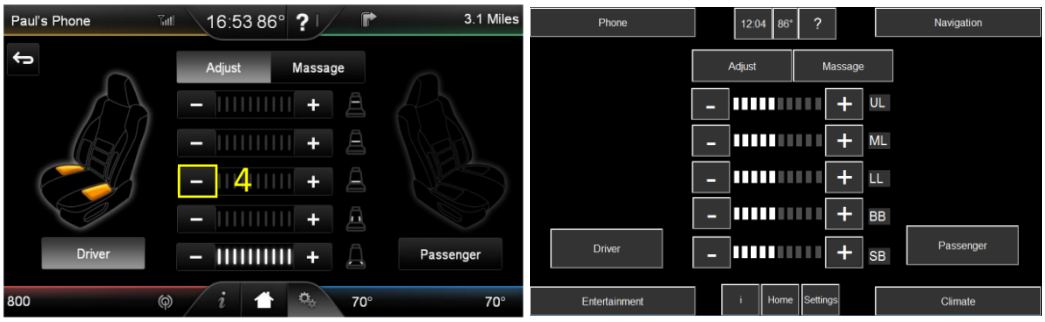
As shown in Figure 8.2, a digital mockup of the center stack touchscreen device, which is a touchscreen-speech interface (*MyFord Touch*), was developed. The yellow-numbers (i.e., 1 through 4) indicate the order of buttons to be clicked to complete the direct-manual tasks using the center stack touchscreen device.



(a) Main Screen



(b) An intermediate screen from Main Screen to the Seat Comfort Screen



(c) Seat Comfort Screen

Figure 8.2 – Center stack touchscreen (on the left) and its digital mockup (on the right)

8.3.3.2. Dashboard-cluster

A digital mockup for the dashboard-cluster was developed (see Figure 8.3). The yellow-numbers (i.e., 1 through 3) indicate the order of buttons to be selected to complete the indirect-manual tasks. Note that the dashboard-cluster is not a touchscreen, but a display, and the contents on the dashboard-cluster should be selected by the steering wheel-mounted 5-way buttons. On the dashboard-cluster, a visual feedback (e.g., highlighted in white or orange) enables drivers to confirm the current position of the cursor. In the digital mockup, blue colors and white texts were used as a visual feedback.

(a) Main Screen

(b) Seat Comfort Screen

Figure 8.3 – Dashboard-cluster (on the left) and its digital mockup (on the right)

8.3.3.3. Steering wheel-mounted 5-way buttons

A digital mockup of the 5-way buttons mounted on the steering wheel was developed (see Figure 8.4). These 5-way buttons are remotely connected with the dashboard-cluster to allow users to conduct indirect-manual tasks.

Figure 8.4 – Five-way-button device (on the left) and its digital mockup (on the right)

8.4. Model Evaluation

Task completion time and workload collected in Chen et al. (2016) were used to validate the model developed in this chapter. Chen et al. (2016) conducted an experimental study using 35 participants (15 females and 20 males; all native English speakers) to evaluate three different methods of seat comfort adjustment including voice, center stack touchscreen, and five-way push buttons mounted on the steering wheel along with a small display on the dashboard-cluster (see Figure 8.5). Of the original experimental outcomes, the experimental data from the task, "*Deflate*

lower lumbar from 50 to 40", were used to evaluate the model. The participants were instructed to adjust the firmness of a particular area of the seat from the level of 50 to 40, using voice-activated commands (i.e., a voice control), a center stack touchscreen (i.e., a direct-manual control), and 5-way buttons and a dashboard-cluster (i.e., an indirect-manual control). In Chen et al. (2016), workload was measured by NASA's task load index (TLX) including mental demand, physical demand, temporal demand, performance, effort, frustration. For the model evaluation, the average of the values from the NASA TLX's six categories was compared to the model's estimated workload based on the QN's server utilization (Wu & Liu, 2007).

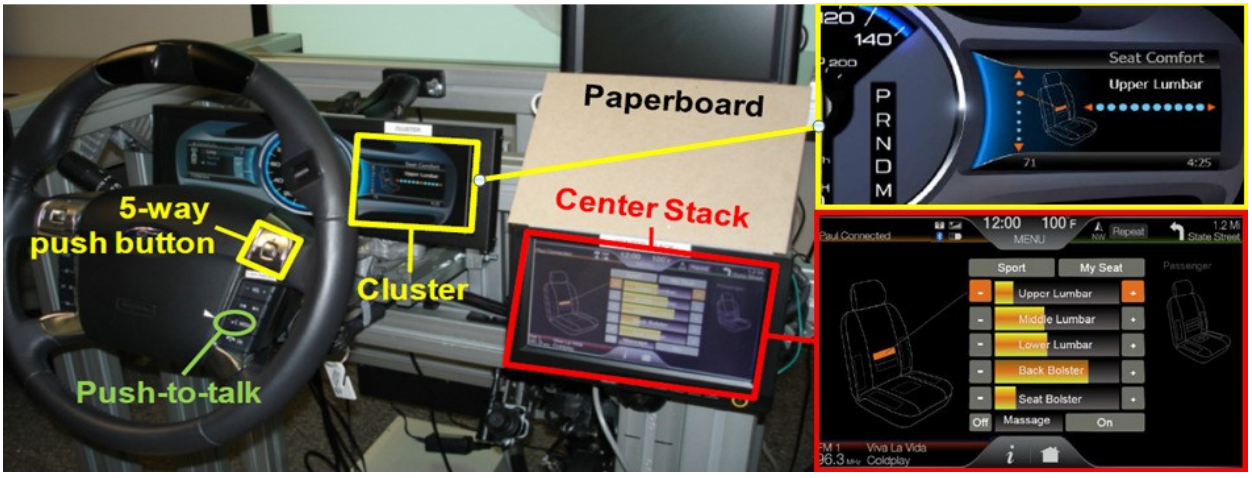


Figure 8.5 – In-vehicle devices for three methods of seat comfort adjustment (Chen et al., 2016)

Ten model simulations were run and processing times were recorded every 50 msec. Waiting time (2 seconds), referring to the time when the user is waiting due to the system's response (Kieras, 1994), was added to the task completion time whenever a voice command was used. The amount of waiting time was determined by the average system response time obtained through video analysis. The model was able to generate results similar to human experimental data. For the task completion time (see Figure 8.6 (a)), the R-square is 88.1%, the root-mean-square error (RMSE) is 2.40 s. For workload (see Figure 8.6 (b)), the R-square is 68.2%, the root-mean-square error (RMSE) is 1.46.

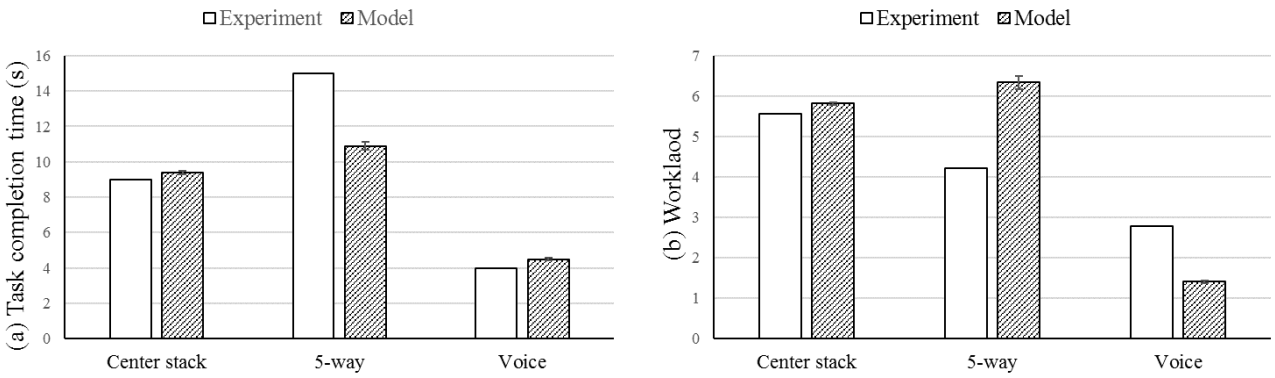


Figure 8.6 – Modeling results of (a) task completion time and (b) workload in comparison to experimental results

8.5. Discussion

This chapter presents a queueing network-based computational model for three different input methods. The model was validated with data from an experimental study, producing similar results in task completion time and workload. Beyond the previous QN-MHP models using a simple voice-related operator (i.e., syllable-level: Jeong & Liu, 2017b), the current model was able to predict the task completion time for a wider range of voice commands (i.e., sentences and words including more than one syllable), by using the pre-determined audio module. Furthermore, visual feedback using colors and texts was implemented to model human indirect-manual controls.

Future research includes expanding the modeling scope, such as developing the predictive models to predict driving performance (e.g., standard deviation of lane position, steering wheel angle) for in-vehicle direct/indirect manual and voice interactions while driving. In addition to the task completion time and workload, I plan to predict other quantitative measures, such as the percentage of eyes-off-road time and voice recognition rates.

According to the literature, age is a critical factor that affects human performance for direct and indirect devices (Roger et al. 2005, Charness, Holley, Feddon, & Jastrzembski, 2004; Murata & Iwase, 2005; Murata & Moriwaka, 2007). Wu and Liu (2007) found that the QN-MHP

architecture could account for the age differences in mental workload and performance between young and old subjects, by considering an aging factor in the servers' processing times. In this regard, I plan to conduct additional modeling research to investigate how users' age difference can affect the performance of in-vehicle direct/indirect manual and voice input tasks.

References

Charness, N., Holley, P., Feddon, J., & Jastrzembski, T. (2004). Light Pen Use and Practice Minimize Age and Hand Performance Differences in Pointing Tasks. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(3), 373–384.

Chen, Y., Tonshal, B., Rankin, J., & Feng, F. (2016). Development of an Integrated Simulation System for Design of Speech-Centric Multimodal Human-Machine Interfaces in an Automotive Cockpit Environment. In *Volume 1A: 36th Computers and Information in Engineering Conference* (p. V01AT02A004). ASME.

Dumas, J. S., & Redish, J. (1999). *A practical guide to usability testing*. Intellect books.

Feng, F., Liu, Y., & Chen, Y. (2017). A computer-aided usability testing tool for in-vehicle infotainment systems. *Computers & Industrial Engineering*, *109*, 313–324.

Feng, F., Liu, Y., Chen, Y., Filev, D., & To, C. (2014, September). Computer-aided usability evaluation of in-vehicle infotainment systems. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, No. 1, pp. 2285-2289). Sage CA: Los Angeles, CA: SAGE Publications.

Green, P. (2000, November). Crashes induced by driver information systems and what can be done to reduce them. In *SAE Conference Proceedings* (pp. 27-36). SAE; 1999.

Harvey, C., & Stanton, N. A. (2013). Modelling the hare and the tortoise: predicting the range of in-vehicle task times using critical path analysis. *Ergonomics*, *56*(1), 16–33.

Harvey, C., Stanton, N. A., Pickering, C. A., McDonald, M., & Zheng, P. (2011a). In-Vehicle Information Systems to Meet the Needs of Drivers. *International Journal of Human-Computer Interaction*, *27*(6), 505–522.

Harvey, C., Stanton, N. A., Pickering, C. A., McDonald, M., & Zheng, P. (2011b). A usability evaluation toolkit for In-Vehicle Information Systems (IVISs). *Applied Ergonomics*, *42*(4), 563–574.

Jacob, R. J. K. (1996). Human-computer interaction: input devices. *ACM Computing Surveys*, *28*(1), 177–179.

Jeong, H., & Liu, Y. (2016, September). Computational Modeling of Finger Swipe Gestures on Touchscreen: Application of Fitts' Law in 3D Space. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 60, No. 1, pp. 1721-1725). Sage CA: Los Angeles, CA: SAGE Publications.

Jeong, H., & Liu, Y. (2017a). Development of a Computational Cognitive Model for In-vehicle Speech Interfaces. In *Proceedings of the 67th Annual Conference and Expo of the Institute of Industrial and Systems Engineers (IISE)*, Pittsburgh, PA, 2171-2176.

Jeong, H., & Liu, Y. (2017b). Modeling of stimulus-response secondary tasks with different modalities while driving in a computational cognitive architecture. In *Proceedings of the Ninth International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Manchester Village, VT, 58-64.

Jeong, H., & Liu, Y. (2017c). Modeling Touchscreen Pinch and Spread Gestures in a Computational Cognitive Architecture. In *Proceedings of the 67th Annual Conference and Expo of the Institute of Industrial and Systems Engineers (IISE)*, Pittsburgh, PA, 2177-2182.

John, B. E., (1990, March). Extensions of GOMS analyses to expert performance requiring perception of dynamic visual and auditory information. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 107-116). ACM.

John, B. E., & Salvucci, D. D. (2005). Multipurpose Prototypes for Assessing User Interfaces in Pervasive Computing Systems. *IEEE Pervasive Computing*, *4*(4), 27–34.

Kieras, D. (1994). A guide to GOMS task analysis. *University of Michigan*.

Klauer, S.G., Dingus, T.A., Neale, V.L., Sudweeks, J.D., & Ramsey, D.J., (2006). The Impact of Driver Inattention on Near-crash/Crash Risk: An Analysis Using the 100-car Naturalistic Driving Study Data (No. DOT HS 810 594). *Virginia Tech Transportation Institute*.

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, V. E.-W., & Green, P. A. (2013). Development and Evaluation of Automotive Speech Interfaces: Useful Information from the Human Factors and the Related Literature. *International Journal of Vehicular Technology*, *2013*, 1–13.

McLaughlin, A. C., Rogers, W. A., & Fisk, A. D. (2009). Using Direct and Indirect Input Devices: Attention Demands and Age-Related Differences. *ACM Transactions on Computer-Human Interaction : A Publication of the Association for Computing Machinery*, *16*(1), 1–15.

Murata, A., & Iwase, H. (2005). Usability of Touch-Panel Interfaces for Older Adults. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *47*(4), 767–776.

Murata, A., & Moriwaka, M. (2007). Applicability of location compatibility to the arrangement of display and control in human – vehicle systems: Comparison between young and older adults. *Ergonomics*, *50*(1), 99–111.

Nielsen, J. (1993), *Usability Engineering*. (San Diego: AP Professional, Academic Press)

Ocak, N., & Cagiltay, K. (2017). Comparison of Cognitive Modeling and User Performance Analysis for Touch Screen Mobile Interface Design. *International Journal of Human–Computer Interaction*, 1–9.

Pettitt, M., Burnett, G., & Stevens, A. (2007). An extended keystroke level model (KLM) for predicting the visual demand of in-vehicle information systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '07* (pp. 1515–1524). New York, New York, USA: ACM Press.

Purucker, C., Naujoks, F., Prill, A., & Neukum, A. (2017). Evaluating distraction of in-vehicle information systems while driving by predicting total eyes-off-road times with keystroke level modeling. *Applied Ergonomics*, *58*, 543-554.

Ranney, T. A., Mazzae, E., Garrott, R., & Goodman, M. J. (2000, July). NHTSA driver distraction research: Past, present, and future. In *Driver distraction internet forum* (Vol. 2000).

Rogers, W. A., Fisk, A. D., McLaughlin, A. C., & Pak, R. (2005). Touch a Screen or Turn a Knob: Choosing the Best Device for the Job. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *47*(2), 271–288.

Salvucci, D. D. (2009). Rapid prototyping and evaluation of in-vehicle interfaces. *ACM Transactions on Computer-Human Interaction*, *16*(2), 1–33.

Schneegaß, S., Pfleging, B., Kern, D., & Schmidt, A. (2011, November). Support for modeling interaction with automotive user interfaces. In *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 71-78). ACM.

Shneiderman, B. (2010). *Designing the user interface: strategies for effective human-computer interaction*. Pearson Education India.

Shutko, J., Mayer, K., Laansoo, E., & Tijerina, L. (2009). *Driver workload effects of cell phone, music player, and text messaging tasks with the Ford SYNC voice interface versus handheld visual-manual interfaces* (No. 2009-01-0786). SAE Technical Paper.

Shutko, J., & Tijerina, L. (2011). Ford's Approach to Managing Driver Attention: SYNC and MyFord Touch. *Ergonomics in Design: The Quarterly of Human Factors Applications*, *19*(4), 13–16.

Stevens, A., Quimby, A., Board, A., Kersloot, T., & Burns, P. (2002). Design guidelines for safety of in-vehicle information systems. *Transportation Research Laboratory*, PA, 3721/01.

Tsimhoni, O., Smith, D., & Green, P. (2004). Address entry while driving: Speech recognition versus a touch-screen keyboard. *Human factors: The Journal of the Human Factors and Ergonomics Society*, *46*(4), 600-610.

Wu, C., & Liu, Y. (2007). Queuing network modeling of driver workload and performance. *IEEE Transactions on Intelligent Transportation Systems*, *8*(3), 528-537.

Chapter 9.

Conclusion and Future Research

9.1. Conclusion

Common methods of testing human performance and usability of certain tasks rely on experimental settings using human subjects. While these methods could provide useful outcomes, it is essential to understand human mental processes in order to effectively improve safety and avoid mental overload. In this regard, I have combined computational cognitive modeling and experimental methods to study mental processes and identify differences in human performance/workload in various conditions, through my dissertation research.

Overall, this dissertation research contributes to our understanding of multimodal multi-task interactions in human-machine systems in (1) touchscreen gestures, (2) audio/speech interaction, and (3) driving control, using both computational cognitive modeling and experimental methods. This is useful especially for human factors engineers and system designers to comprehend human multitasking behavior mechanisms so that they can design and evaluate the appropriate systems for enhancing human safety and performance.

Successfully developed computational models using the QN-MHP architecture helped produce performance and workload predictions. With minimum manpower resources, the models can account for human performance and workload. In other words, these models provided significant contributions to reduce the time and costs needed at the early stage, such as creating prototypes and conducting usability tests using human subjects. Using multiple validation

measures, the models were evaluated by experimental studies and improved so eventually, the models were able to provide more solid and reliable predictions.

9.2. Summary of each chapter

- Chapter 2: An experimental study was conducted to investigate how finger-touch input performance (i.e., task completion time, failure status, and error rate) and subjective ratings (i.e., performance and physical demand) are influenced by touchscreen gestures' type and direction. The results showed that swipe was approximately 4.5 times faster than drag, but pinch and spread showed no significant difference in task completion time. Dragging and pinching showed more failures or higher error rates compared to swiping and spreading, respectively. One-touch gestures in the horizontal directions were rated to have higher performance and lower physical demand than those in the vertical and diagonal directions. Two-touch gestures in the horizontal directions took the shortest time but caused more failures and higher error rates.
- Chapter 3: A computational model was developed to predict finger-drag gesture performance on touchscreen devices, by integrating the queueing network cognitive architecture and three-dimensional motion tracking. Specifically, the QN-based model was developed to predict two execution times: finger movement time of drag-gesture (i.e., only the motion time of the finger touched and dragged on the surface of the touchscreen) and a comprehensive process time of drag-gesture (i.e., the entire process time to complete the finger-drag task). To develop predictive models for the finger movement time of drag gesture, eleven participants' motion data were collected and a regression analysis with parameters of hand-finger anthropometric data and eight

angular directions was conducted. Human subject data of Jeong and Liu (2017a) were used to evaluate the QN-based model, generating similar outputs (R^2 is more than 80% and RMS is less than 300 msec) of the execution times.

- Chapter 4: This chapter reported a driving simulation study of the impacts of both secondary task modality (4 levels) and road curvature (2 levels) on driver behavior. Eye movements, lane-keeping performance, and subjective workload of 24 participants were measured. The results showed that drivers performing secondary tasks using visual stimuli or manual responses on curved roads fixated less frequently and with shorter durations on the road and showed poorer lane-keeping performance compared to other modalities. In addition, when driving on sharper curves with a secondary task, drivers looked at the road more frequently and longer, but their lane-keeping performance was lower (i.e., higher standard deviations of lane position and of steering wheel angle). Participants reported higher visual demand when performing visual-speech types of tasks compared to auditory-manual types of tasks.
- Chapter 5: This chapter introduced a computational human performance model based on the queueing network cognitive architecture to predict driver's eye glances and workload for four stimulus-response secondary tasks (i.e., auditory-manual, auditory-speech, visual-manual, and visual-speech types) while driving. The model was evaluated with the empirical data from 24 subjects, and the percentage of eyes-off-road time and driver workload generated by the model were similar to the human subject data.
- Chapter 6: The study reported in this chapter aimed to examine the effects of road geometry (i.e., road curvature and curve direction) and lead vehicle on horizontal curve driving performance. Twenty-four participants performed two driving tasks (i.e., car-following and free-flow conditions) in simulated driving scenarios including curved roads in both right and

left directions to measure their driving performance (e.g., speed, lane position, steering wheel angle, and time and distance headways). The results showed that road curvature affected driving performance, especially in mean speed, the variability of lane-keeping performance, and headway. Moreover, the mean speed was affected by both road curvature and whether a lead vehicle existed. Findings from this study provide empirical data that can be used for driving safety on horizontal, curved roads.

- Chapter 7: To understand and predict driver performance on curved roads, a computational model was developed in a cognitive architecture, the QN-MHP, with the integration of vehicle dynamics principles (i.e., how to steer based on near and far angles) and the reference trajectory tracking method (i.e., how to steer on the road varying with radius of road curvature). The model was implemented with four major components: road information, vehicle dynamics, visual perception, and cognition & motor controls. The model outputs were validated with the corresponding human subject performance in the literature. The performance results of the model highly fitted the human subject data such as steering wheel angle.
- Chapter 8: While many human subject studies have been conducted to evaluate in-vehicle manual and speech systems, there are few modeling studies using cognitive architectures and quantitative prediction methods. This chapter presented a computational cognitive model for three different input methods: direct-manual control (using touchscreen buttons), indirect-manual control (using steering-mounted buttons paired with a small display in the dashboard-cluster), and voice control. The model was evaluated with data from 35 human subjects in a laboratory experiment using in-vehicle interfaces to adjust the comfort level of the driver

seat. The model produced results similar to actual human performance (i.e., task completion time, workload).

9.3. Future research

9.3.1. Modeling human performance in other domains

Currently, the models introduced in this dissertation were mainly on simulating human-machine interactions while using touchscreen devices and speech interfaces, and driving vehicles. Considering the high compatibility of the QN-MHP architecture with other domains, it will be possible to expand the scope and application of cognitive modeling. Future research will be to extend my QN-MHP computational modeling skills learned through this dissertation to other human-machine system domains such as pilot–aircraft, operator–nuclear power system, doctor–medical device, and human-robot system. The suggested research will be promising in terms of the extensibility to these areas in human factors society.

9.3.2. Modeling human performance for individual differences

The models developed through this dissertation simulated the "normal and average" human behavior, although there exists a wide range of spectrums of the human beings in the real world. It would be possible to investigate individual differences in human performance, such as by age, gender, anthropometry, trained level, and disabled level, by applying relevant psychology and neuroscience findings from the literature to the functions of the QN-MHP architecture.

9.3.3. Modeling other measures of human performance

The current models in this dissertation were able to predict general performance measures, such as task completion time, workload, eye movement (eyes-off-road time), and driving performance (mean steering wheel angle). However, the current models are lack of some

performance outcomes, such as error rate in touchscreen finger-input tasks, the error rate in voice input tasks, and a variety of driving performance indexes (e.g., speed, lane position, distance and time headways). Future studies are to enrich the scope of the model outcomes so that the model can predict a wider range of human performance in complex and specific tasks.

9.3.4. Modeling human performance in combination with other technologies

The current models in this dissertation were successfully integrated with relevant technologies in order to develop new operators for the QN-MHP models, such as three-dimensional tracking (for touchscreen-related operators in Chapter 4), vehicle dynamics and reference trajectory tracking (for steering-wheel- and pedal-related operators in Chapter 7). Going forward, further research is to use other technologies for the new operators in the QN-MHP multimodal interaction models. For instance, tactile sensory-related operators can be developed, using haptic (or tactile) technology. Moreover, eye-tracking pattern-related operators can be developed, using computer vision technology that enables identifying visual cues associated with humans' decision-making.