MODELING OF STIMULUS-RESPONSE SECONDARY TASKS WITH DIFFERENT MODALITIES WHILE DRIVING IN A COMPUTATIONAL COGNITIVE ARCHITECTURE

Heejin Jeong[1], Yili Liu[1]
[1]University of Michigan
Ann Arbor, Michigan, USA
Email: heejinj@umich.edu

Summary: This paper introduces a computational human performance model based upon 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. Future studies aim to extend the types of voice announcements/commands to enable Human-Machine-Interface (HMI) evaluations with a wider range of usability test for in-vehicle infotainment system developments.

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 were 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 paper, we 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) 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 of 24 participants.

MODEL DEVELOPMENT

As shown in Figure 1, the QN-MHP architecture consists of three subnetworks (perception, cognition, and motor). In the architecture, it is assumed that each subnetwork includes multiple servers (1-8; A-G; W-Z) and each server has its own function, based on the findings from previous psychology and neuroscience studies (Liu, Feyen, & Tsimhoni, 2006).

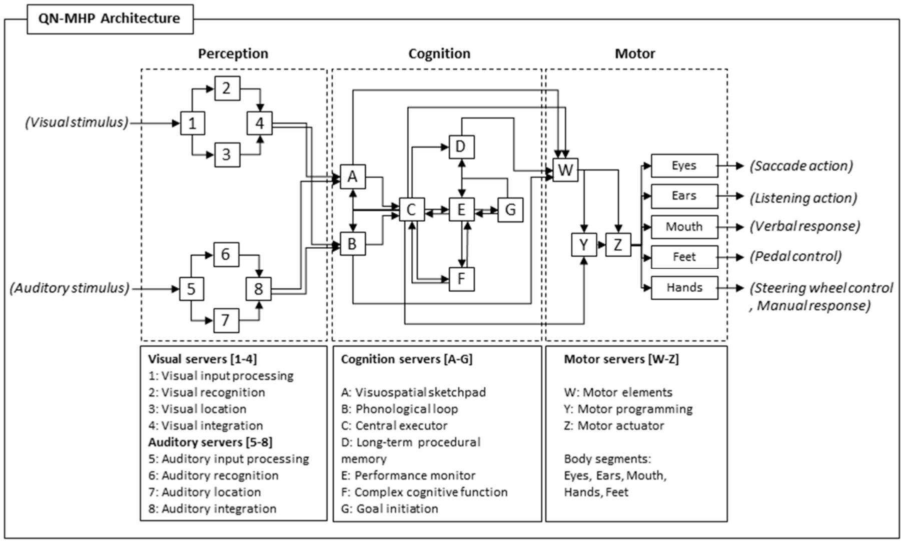

Figure 1. QN-MHP architecture (Liu et al., 2006)

To implement human performance models for particular tasks, using the QN-MHP architecture, it is required to (1) analyze the tasks, (2) develop necessary operators based upon result of the task analysis, and (3) develop digital device mockups, especially for HMI evaluations. The operators, referring to the most elementary components of the task, are developed with related experimental research findings and quantitative models.

Task Analyses

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). 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 1, showing the result of the task analyses, "*Look-at*", "*Listen-to*", "*Click-with-finger*" are the examples of operator, whereas "<target type>", "<device id>", "<*x*, *y*>" are the examples of parameter.

Table 1. NGOMSL-style descriptions of the four stimulus-response tasks

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

[A-S task]	[V-S task]
Goal: Listen to a number and say the number	Goal: Look at a number and say the number
TC 1: *Listen-to* <target value>	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 2: *Store* the <target value> on 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: *Compare* <target value> to <expected value>
If match, return result =1, else return result = 0	If match, return result =1, else return result = 0
TC 5: *Decide*	TC 5: *Decide*
If result = 1, go to TC 6, else go to TC 1	If result = 1, go to TC 6, else go to TC 1
TC 6: *Say <a number>*	TC 6: *Say <a number>*
TC 7: Return with goal accomplished	TC 7: Return with goal accomplished

Development of Operators

Because the recent QN-MHP-based models had 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, we 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 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. 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 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 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 upon the estimation of how far/long the hand reaches to 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:

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

, where *a* and *b* are empirical regression coefficients, varying in 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, we assumed that the model uses a right hand. Also, we 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 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.

Development of Digital Device Mockups

Using MATLAB Graphical User Interface Design Environment (GUIDE), digital device mockups of the NBack App were developed. Figure 2 shows the digital mockup of the V-M task, as an example. Figure 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 2-(b). Where the model looks at and clicks on the digital mockup are indicated by yellow-hatched and white-dotted squares, respectively.

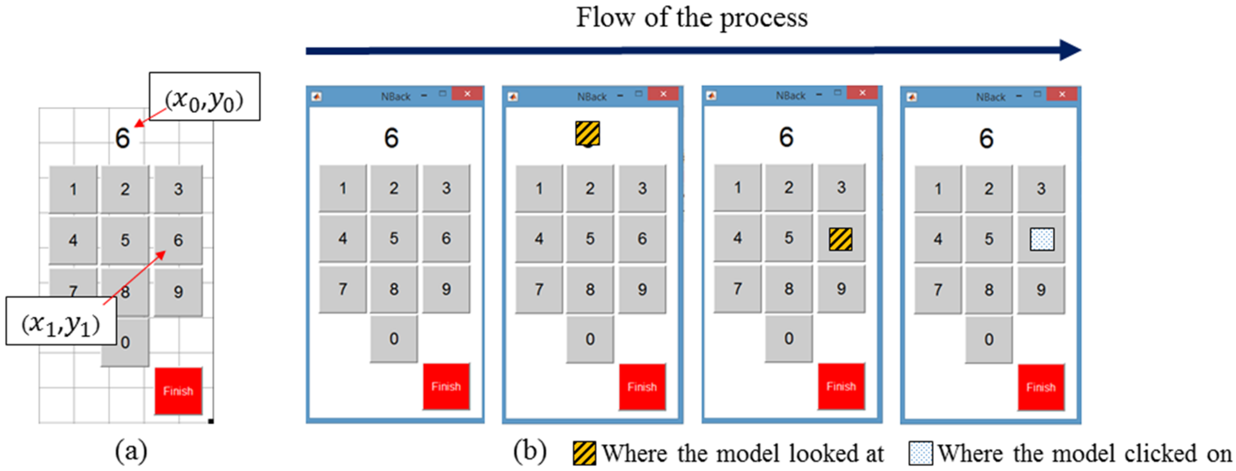

Figure 2. (a) An example of digital mockup for the NBack App and (b) its usage for a visual-manual secondary task while driving

MODEL VALIDATION

Experiment 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 effect 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 and 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) were collected every 50 msec. As shown in Figure 3, the model was able to generate quite similar results of both the percentage of eyes-off-road time ($R^2 = 0.88$, RMS = 4.95) and workload ($R^2 = 0.99$, RMS = 1.16) to the human subject data.

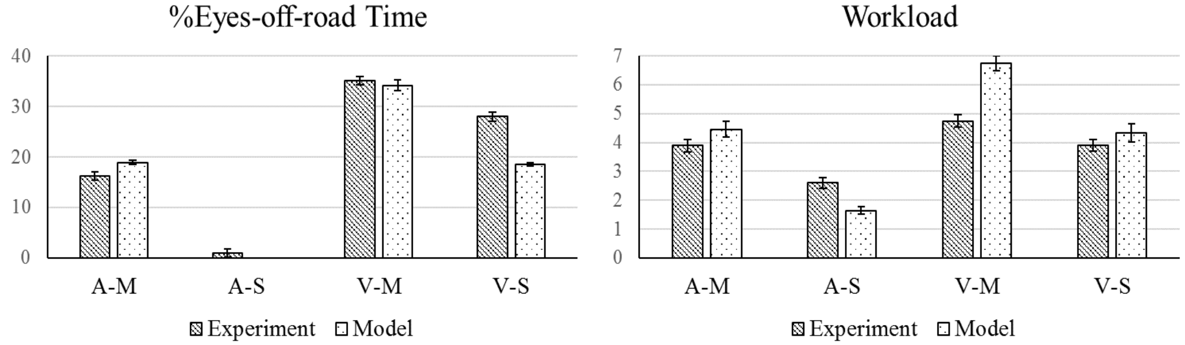

Figure 3. Modeling results of %Eyes-off-road time and workload in comparison to human results

CONCLUSION AND DISCUSSION

This paper 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, we 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 we 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).

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).