Modeling the Effects of Auditory Display Takeover Requests on Drivers' Behavior in Autonomous Vehicles

Sangjin Ko
Virginia Tech
Blacksgurg, VA 24061, USA
sangjinko@vt.edu

Yiqi Zhang
Pennsylvania State University
University Park, PA 16802, USA
yuz450@psu.edu

Myounghoon Jeon
Virginia Tech
Blacksgurg, VA 24061, USA
myounghoonjeon@vt.edu

AutomotiveUI '19 Adjunct, September 21–25, 2019, Utrecht, Netherlands

ACM ISBN 978-1-4503-6920-6/19/09.
https://doi.org/10.1145/3349263.3351508

Abstract

In semi-autonomous vehicles (SAE level 3) that require driver's engagement in critical situations, it is important to secure reliable control transitions. There have been many studies on investigating appropriate auditory displays for takeover request (TOR) but most of them were empirical experiments. In the present study, we established two computational models using a Queuing Network Model Human Processor (QN-MHP) framework to predict a driver's reaction time to auditory displays for TOR. The reaction time for different sound types were modeled based on the results of subjective questionnaire in empirical studies. Separately, the reaction times for various non-speech sounds were modeled by using acoustical characteristics of sounds and previous empirical studies. It is one of a few attempts modeling the effects of auditory displays for TOR on the reaction time in autonomous driving. The current study will contribute to driving research by allowing us to simulate and predict drivers' behavior.

Author Keywords

Auditory warning; takeover requests; computational model; queuing network model human processor (QN-MHP); driver behavior

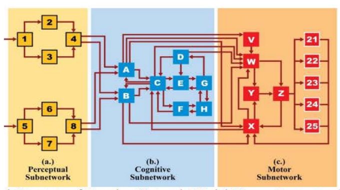

Figure 1: The general structure of the queuing network-model human processor (QN-MHP). It consists of 3 sub-networks (perceptual, cognitive, and motor subnetworks) and 26 servers (square boxes in each sub-network).

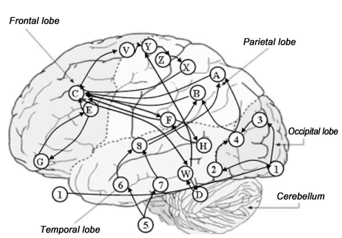

Figure 2: Approximate mapping of servers in the queuing network model onto the human brain

CCS Concepts

• Human-centered computing~HCI theory, concepts and models

Introduction

There has been considerable research on how to ensure a safe transition from autonomous to manual driving in semi-autonomous vehicles. This is because semi-autonomous vehicles still require drivers' monitoring to take control of the vehicles in unexpected situations [1, 2]. In line with this, there have been experimental studies on driver performance in response to takeover requests in highly autonomous vehicles [3-6]. Together with visual displays, auditory displays have been widely investigated based on multiple resource theory (MRT) [7, 8] that predicts better performance when different modalities are employed. These studies have evaluated the difference in performance between different types of auditory displays such as speech, auditory icons, earcons, and spearcons [9-12]. However, most studies were conducted through empirical human-subject studies requiring extensive efforts and resources. This type of research again requests a new study to apply to other contexts. To tackle this issue, there have been attempts to model performances and workload on multitasks [13-15] and driver performance on auditory displays [16-18]. In the same line, we established new cognitive models to predict driver performance in autonomous vehicles by using the Queuing Network-Model Human Processor framework (QN-MHP) [2]. The new models developed using subjective measures and empirical data sets and included different types of auditory displays for takeover requests in the autonomous driving context. In this paper, the concept of QN-MHP, previous cognitive models, current modeling works, and future works are presented.

Related Works

Cognitive Modeling

To estimate the cognitive process and behaviors on the stimuli, various cognitive models have been used, such as MHP (Model Human Processor), GOMS (Golas, Operators, Methods, and Selection rules), EPIC (Executive-Process/Interactive Control), ACT-R (Adaptive Control of Thought Rational) and SOAR (State, Operator, And Result) [19-22]. MHP is based on the hypothesis that the mind acts as a processor of information like a computer. In the lines of human information processing theory, MHP has three stages including Perception, Cognition, and Motor. Information signals pass through each stage and can move iteratively through each stage [22]. This model is an approximating cognitive model that permits approximate calculations [19]. In the present study, we applied the QN-MHP framework which is a computational cognitive architecture based on MHP.

QN-MHP Framework

The QN-MHP framework represents the human cognition system as a queuing network-based on several similarities to brain activities [15, 23, 24]. QN-MHP consists of three subnetworks: perceptual, cognitive, and motor subnetworks, as described in Figure 1 [15]. Brain regions with similar functions can be represented as servers and neural pathways connecting them as routes in the queuing network, as described in Figure 2.

The queuing network model divides each stage into a subnetwork of a small number of servers and thus, has a level of granularity that falls between the neural network and MHP model. QN-MHP has been developed as an integration of queuing networks and MHP for both mathematical modeling and real-time generation of

Figure 3. 2048 Game as a secondary task

Figure 4. Center console with a visual display

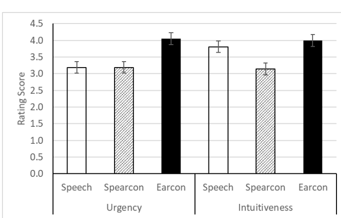

Figure 5. Results of subjective questionnaires in the experiment

psychological behavior. Each subnetwork is constructed of multiple servers and links among these servers. Each server is an abstraction of a brain area with specific functions, and links among servers represent neural pathways among functional brain areas. The neurological processing of stimuli is illustrated in the transformation of entities passing through routes in QN-MHP.

Previous models using QN-MHP

QN-MHP has been successfully used to generate human performance and mental workload in real time. They included driver performance and workload, transcription typing performance, and visual–manual tracking performance and mental workload [13, 14, 25]. In the current paper, we enhanced the previous QN-MHP model that accounts for the effect of loudness and semantics of speech warning messages on human responses. The previous work was done specifically to include modeling collision avoidance control performance and warning response type selection and execution [16-18]. The present paper enhanced the model to predict the reaction time of a driver to takeover request signals in an autonomous vehicle while playing a distracting game.

Modeling Works

Experiment

Two experiments were conducted with the same experimental design setting but with different auditory displays [26, 27]. The experiments applied a within-subjects design with different auditory warnings for takeover requests. The reaction times to takeover of the control were measured as dependent variables. Also, as subjective measures, NASA-TLX and subjective rating scores regarding user experience were measured. The driving scenarios were set in a rural area with the car driving on a two-lane road without oncoming traffic and intersection. The speed limit was set at 50 miles per hour. Three different situations with takeover requests were presented to each participant. These situations on the road that triggered the participant to take over included a deer, a parked car, and a service vehicle, all blocking most or all of the driving lanes. During the autonomous driving mode, the participants played the online game, 2048 (Figure 3) on a laptop placed next to them on the center console (Figure 4). They were instructed to focus on playing the game during the autonomous mode. After around one minute, one of the auditory displays was presented for the participants to take over control of the vehicle due to a situation on the road.

Modeling 1: Different Sound Types

In the first modeling work, three different types of auditory displays for takeover requests were employed [26], including speech, spearcon [28], and earcon [29]. The results of subjective questionnaires were depicted in Figure 5. The previous studies on trust in the system showed that the increase in accuracy and intelligibility led to the increase in trust [30-34]. Following them, the probability of traveling via a shorter route in the previous QN-MHP model [16-18] influenced by perceived trust of the warning, P_{tr}, was manipulated. To be specific, P_{tr} was modeled by warning reliability *wr* and intuitiveness of auditory warnings I_{ws} as follows.

$$P_{tr} = \frac{1}{2}\left(wr + I(ws)\right)$$

Likewise, the probability influenced by perceived urgency of warnings, P_{wu}, was modeled by warning loudness *wl* and urgency level *U(ws)* as follows.

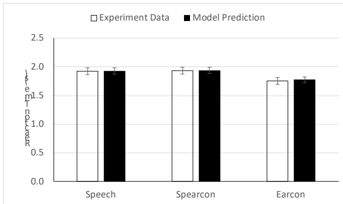

Figure 6. Reaction times in the experiment and Model 1

	Hyundai	Tesla	New
A*	1498	553	530
B*	7.5	1.62	0.45
C*	1	1	5

Table 1: Acoustic characteristics of three display types
**A: Fundamental frequency (Hz)*
**B: Repetitions/sec*
**C: Number of Dominant frequency*

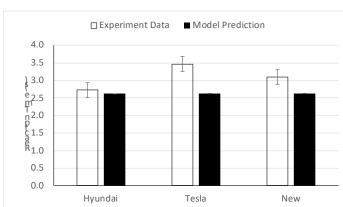

Figure 7. Reaction times in the experiment and Model 2

$$P_{wu} = \frac{1}{2}\,(U(wl) + U(ws))$$

Consequently, the reaction time for takeover requests, T_r, is modeled by the stimulus processing time of a route and the probability of a stimulus travelling through it, denoted by the equation below.

$$T_r = \sum_{u=1}^{2} PT_u \times P_u(i, j, wl, wsm, wt, ws, wr)$$

wt is the warning lead time and *ws* denotes warning style. PT_u is defined as the processing time for a stimulus through a route *u*.
The model prediction for the reaction time for takeover requests is shown in Figure 6. The model prediction was validated with the experimental data set via the Pearson correlation coefficient (R-squared) and the root mean-squared error (RMSE). The RMSE was 0.073 seconds (73ms) with an R-square of 0.925.

Modeling 2: Different Non-speech Sounds
In the second modeling work, three different types of non-speech sounds (two already implemented sounds and one newly designed sound) were employed [27]. At this time, perceived urgency that affects the probability of choosing a shorter route, P_{wu}, was estimated based on acoustic characteristics. Representative characteristics of three non-speech sounds were depicted in Table 1. To be specific, the new parameter, U(wac), was added to account the perceived urgency from acoustic characteristics of sounds. P_{wu} was modeled as below.

$$P_{wu} = \frac{1}{3}\,(U(wl) + U(wsm) + U(wac))$$

Specifically, the values of U(wac) was estimated as below.

$$U(wac) = \frac{1}{3}\,(U(wfreq) + U(wrep) + U(wpit))$$

Each term represents the acoustic characteristic that affects perceived urgency; *U(wfreq)* for the fundamental frequency, *U(wrep)* for the number of the repetitions per second, and *U(wpit)* for the pitch range of dominant frequencies. Coefficients in each term were derived from the previous studies showing linear relationships between each aspect and perceived urgency.
The model predictions and the compared data are shown in Figure 7. The RMSE was 0.562 seconds (562ms) with an R-square of 0.958. However, the differences between the highest and the lowest predictions were smaller than validation data.

Conclusion and Future Work

In summary, the presented models reliably predicted a driver's reaction times in response to different types of auditory takeover requests in autonomous vehicles. The models will be utilized to further investigate and elaborate auditory display designs in autonomous vehicles by providing the cost-effective way. Researchers can simulate a driver's behavior with various factors of auditory displays. The models can be expanded by spanning varying parameters in different contexts. Eventually, it will contribute to establishing design guidelines for an international standard of takeover request warnings for autonomous vehicles.

References

[1] SAE, T., "Definitions for terms related to driving automation systems for on-road motor vehicles," *SAE Standard J,* vol. 3016, 2016.

[2] Tsimhoni, O. and Liu, Y., "Modeling steering using the queueing network—model human processor (QN-MHP)," in *Proceedings of the human factors and ergonomics society annual meeting*, 2003, vol. 47, no. 16, pp. 1875-1879: SAGE Publications Sage CA: Los Angeles, CA.

[3] Gold, C., Damböck, D., Lorenz, L., and Bengler, K., "“Take over!” How long does it take to get the driver back into the loop?," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2013, vol. 57, no. 1, pp. 1938-1942: SAGE Publications Sage CA: Los Angeles, CA.

[4] Melcher, V., Rauh, S., Diederichs, F., Widlroither, H., and Bauer, W., "Take-over requests for automated driving," *Procedia Manufacturing,* vol. 3, pp. 2867-2873, 2015.

[5] Merat, N., Jamson, A. H., Lai, F. C., and Carsten, O., "Highly automated driving, secondary task performance, and driver state," *Human factors,* vol. 54, no. 5, pp. 762-771, 2012.

[6] Zeeb, K., Buchner, A., and Schrauf, M., "What determines the take-over time? An integrated model approach of driver take-over after automated driving," *Accident Analysis Prevention,* vol. 78, pp. 212-221, 2015.

[7] Wickens, C. D., "Multiple resources and performance prediction," *Theoretical issues in ergonomics science,* vol. 3, no. 2, pp. 159-177, 2002.

[8] Wickens, C. D., "Multiple resources and mental workload," *Human factors,* vol. 50, no. 3, pp. 449-455, 2008.

[9] Bazilinskyy, P. and de Winter, J., "Auditory interfaces in automated driving: an international survey," *PeerJ Computer Science,* vol. 1, p. e13, 2015.

[10] McKeown, D., "Candidates for within-vehicle auditory displays," 2005: Georgia Institute of Technology.

[11] Richie, E., Offer-Westort, T., Shankar, R., and Jeon, M., "Auditory Displays for Take-Over in Semi-automated Vehicles," in *International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management*, 2018, pp. 623-634: Springer.

[12] Sabic, E., Mishler, S., Chen, J., and Hu, B., "Recognition of Car Warnings: An Analysis of Various Alert Types," in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2017, pp. 2010-2016: ACM.

[13] Jeong, H. and Liu, Y., "Modeling of stimulus-response secondary tasks with different modalities while driving in a computational cognitive architecture," 2017.

[14] Salvucci, D. D., "Modeling driver behavior in a cognitive architecture," *Human factors,* vol. 48, no. 2, pp. 362-380, 2006.

[15] Wu, C. and Liu, Y., "Queuing network modeling of driver workload and performance," *IEEE Transactions on Intelligent Transportation Systems,* vol. 8, no. 3, pp. 528-537, 2007.

[16] Zhang, Y., Changxu, W., and Jingyan, W., "Mathematical Modeling of the Effects of Speech Warning Characteristics on Human Performance and Its Application in Transportation Cyberphysical Systems," *IEEE Transactions on*

Intelligent Transportation Systems, vol. 17, no. 11, pp. 3062-3074, 2016.

[17] Zhang, Y. and Changxu, W., "Modeling the Effects of Warning Lead Time, Warning Reliability and Warning Style on Human Performance Under Connected Vehicle Settings," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2018, vol. 62, no. 1, pp. 701-701: SAGE Publications Sage CA: Los Angeles, CA.

[18] Zhang, Y., "Modeling the Effects of Warning Lead Time, Warning Reliability and Warning Style on Human Performance Under Connected Vehicle Settings," Ph.D. Dissertation, State University of New York at Buffalo, 2017.

[19] Newell, A. and Card, S. K., "The prospects for psychological science in human-computer interaction," *Human-computer interaction,* vol. 1, no. 3, pp. 209-242, 1985.

[20] Laird, J. E., Newell, A., and Rosenbloom, P. S., "Soar: An architecture for general intelligence," *Artificial intelligence,* vol. 33, no. 1, pp. 1-64, 1987.

[21] Kieras, D. E. and Meyer, D. E., "An overview of the EPIC architecture for cognition and performance with application to human-computer interaction," *Human–Computer Interaction,* vol. 12, no. 4, pp. 391-438, 1997.

[22] Feyen, R. G., "Modeling human performance using the queuing network-model human processor (QN-MHP)," 2002.

[23] Wu, C. and Liu, Y., "Queuing network modeling of the psychological refractory period (PRP)," *Psychol Rev,* vol. 115, no. 4, pp. 913-54, Oct 2008.

[24] Zhang, Y. and Wu, C., "Learn to Integrate Mathematical Models in Human Performance Modeling," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2017, vol. 61, no. 1, pp. 776-780: SAGE Publications Sage CA: Los Angeles, CA.

[25] Salvucci, D. D., "Rapid prototyping and evaluation of in-vehicle interfaces," *ACM Transactions on Computer-Human Interaction,* vol. 16, no. 2, p. 9, 2009.

[26] Jeon, M., "Multimodal Displays for Take-over in Level 3 Automated Vehicles while Playing a Game," in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, p. LBW0229: ACM.

[27] Kutchek, K. and Jeon, M., "Takeover and Handover Requests using Non-Speech Auditory Displays in Semi-Automated Vehicles," in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, p. LBW0228: ACM.

[28] Walker, B. N. *et al.*, "Spearcons (speech-based earcons) improve navigation performance in advanced auditory menus," *Human Factors,* vol. 55, no. 1, pp. 157-182, 2013.

[29] Blattner, M. M., Sumikawa, D. A., and Greenberg, R. M., "Earcons and icons: Their structure and common design principles," *Human–Computer Interaction,* vol. 4, no. 1, pp. 11-44, 1989.

[30] Bonsall, P. W. and Joint, M., "Driver compliance with route guidance advice: the evidence and its implications," in *Vehicle Navigation and Information Systems Conference, 1991*, 1991, vol. 2, pp. 47-59: IEEE.

[31] Fox, J. and Boehm-Davis, D., "Effects of age and congestion information accuracy of advanced traveler information systems on user trust and compliance," *Transportation Research Record:*

Journal of the Transportation Research Board, no. 1621, pp. 43-49, 1998.

[32] Jonsson, I.-M., Harris, H., and Nass, C., "How accurate must an in-car information system be?: consequences of accurate and inaccurate information in cars," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2008, pp. 1665-1674: ACM.

[33] Lee, J. and Moray, N., "Trust, control strategies and allocation of function in human-machine systems," *Ergonomics,* vol. 35, no. 10, pp. 1243-1270, 1992.

[34] Lim, B. Y., "Improving trust in context-aware applications with intelligibility," in *Proceedings of the 12th ACM international conference adjunct papers on Ubiquitous computing-Adjunct*, 2010, pp. 477-480: ACM.