IMPACT OF COGNITIVE WORKLOAD ON PHYSIOLOGICAL AROUSAL AND PERFORMANCE IN YOUNGER AND OLDER DRIVERS

Joonwoo Son[1], Bruce Mehler[2], Taeyoung Lee[1], Yunsuk Park[1],
Joseph Coughlin[2], & Bryan Reimer[2]
[1] Daegu Gyeongbuk Institute of Science & Technology (DGIST), Daegu, South Korea
[2] Massachussets Institute of Technology (MIT) AgeLab & New England University
Transportation Center, Cambridge, Massachusetts, USA
Email: json@dgist.ac.kr

Summary: Two groups, aged 25-35 and 60-69, engaged in 3 levels of a delayed auditory recall task while driving a simulated highway. Heart rate and skin conductance increased with each level of demand, demonstrating that these indices can correctly rank order cognitive workload. Effects were also observed on speed and SD of lane position, but they were subtle, nonlinear, and did not effectively differentiate. Patterns were quite consistent across age groups. These findings on the sensitivity of physiological measures replicate those from an on-road study using a similar protocol. Together, the results support the validity of using these physiological measures of workload in a simulated environment to model differences likely to be present under actual driving conditions.

INTRODUCTION

Physiological measures have been proposed as useful metrics for assessing workload in the user interface design assessment and optimization process (Lenneman & Backs, 2010; Mehler, Reimer, Coughlin & Dusek, 2009). While physiological indices have been used to measure workload for some time, particularly in the aviation literature (Wilson, 2002), until recently data on the sensitivity of such measures in the driving environment has been either limited or contradictory. An early study (Brookhouis, De Vies & De Waard, 1991) showed that a task delivered over a cell phone resulted in an increase in heart rate while driving and another (Brookhouis & De Waard, 2001) showed elevations during heightened driving demands such as entering a traffic circle. However, an ambitious undertaking involving three different European centers (HASTE Project) that compared heart rate, heart rate variability and skin conductance measures for two task types (a surrogate visual interface task and an auditory task) at three levels of difficulty for each and across a fixed-based simulator, moving-base simulator, and under on-road conditions, produced contradictory results in the different environments and generally either no significant differentiation across task difficulty levels or counterintuitive response patterns (Engstrom, Johansson & Ostlund, 2005).

More supportive data have since been presented by Lenneman and Backs (2010) from a simulation study that monitored a range of cardiovascular measures during single task driving and dual task conditions using a relatively easy visual n-back target matching task (0-back) and a much more difficult version (3-back). They found that heart rate was significantly higher during both the 0-back and the 3-back conditions than during single task driving. The difference in heart rate between the 0-back and the 3-back was also statistically significant. Cardiac pre-ejection period (PEP) and a respiratory sinus arrhythmia (RSA) measure were less sensitive, being able to differentiate single task driving from the 3-back task but not the 0-back task. Another simulation

study (Mehler et al., 2009) employed an auditory delayed digit recall task in a variant n-back format with three levels of task difficulty that allowed a finer grain look at the question of sensitivity. The research design increased workload in a sequential order from single task driving through each of the three difficulty levels (0-, 1-, and 2-back) and found that heart rate increased in a statistically significant manner at each level. Skin conductance level (often used as a measure of emotional arousal) also increased significantly from single task driving to 0-back, again from 0-back to the 1-back demand level, and then leveled out during the 2-back. Reimer, Mehler, Coughlin, Godfrey et al. (2009) used the same auditory recall task and presentation order in a modest sized on-road evaluation to assess the generalizability of the simulator findings. They found essentially the same reactivity pattern for heart rate in the field with significant discrimination between single task driving and 0-back, between 0-back and 1-back, and a further but non-significant increase during the 2-back. Skin conductance was again sensitive at indicating initial changes in demand between single task driving and the 0-back but did not provide further discrimination at the higher levels. Each of these studies used predominately young adults.

Mehler, Reimer and Coughlin (2010) extended this work by looking at subjects in their 20s, 40s and 60s to examine whether age impacts the relative sensitivity of heart rate and skin conductance measures under on-road conditions. They also assessed the extent to which the sequential presentation of the n-back tasks from easiest to most difficult influenced the apparent ceiling effect, particularly in skin conductance, at the higher levels of demand. They found that when the three levels of task difficulty were randomly ordered, and a recovery interval was provided between tasks, a near linear increase in heart rate appeared across the demand levels. For both heart rate and skin conductance, each task level was statistically differentiated from single task driving and from each other. These findings demonstrate that both heart rate and skin conductance provide the sensitivity to discriminate incremental changes in cognitive workload induced by this task under actual driving conditions. The study also found that, in relatively healthy individuals, the pattern of change across demand levels was consistent across age groups.

The present paper examines the extent to which the basic findings observed in Mehler et al. (2010) are replicable by another research team in a different setting and culture (Korea vs. USA), with somewhat different age groupings, and in a simulator. Sensitivity in a simulation environment and the question of the extent to which patterns observed in simulator model what happens under actual driving conditions, are critical considerations in assessing the degree to which physiological measures of cognitive workload can be productively used in simulation for interface design evaluation.

METHOD

Subjects

Subjects were required to meet the following criteria: age between 25-35 or 60-69, drive on average more than twice a week, be in self-reported good health and free from major medical conditions, not take medications for psychiatric disorders, score 25 or greater on the mini mental status exam to establish reasonable cognitive capacity and situational awareness, and have not

previously participated in a simulated driving study. The sample consisted of 30 males: 15 in the 25-35 age range (M=27.9, SD=3.13) and 15 in the 60-69 range (M= 63.2, SD= 1.74).

Experimental setup

The experiment was conducted in the DGIST fixed-based driving simulator, which incorporated STISIM Drive™ software and a fixed car cab (see Figure 1). Graphical updates to the virtual environment were computed using STISIM Drive™ based upon inputs recorded from the OEM accelerator, brake and steering wheel which were all augmented with tactile force feedback. The virtual roadway was displayed on a 2.5m by 2.5m wall-mounted screen at a resolution of 1024 x 768. Sensory feedback to the driver was also provided through auditory and kinetic channels. Distance, speed, steering, throttle, and braking inputs were captured at a nominal sampling rate of 30 Hz. Physiological data were collected using a MEDAC System/3 unit and NeuGraph™ software (NeuroDyne Medical Corp., Cambridge, MA). A display was installed on the screen beside the rear-view mirror to provide information about the elapsed time and the distance remaining in the drive. The simulator configuration, STISIM Drive™ protocols, and physiological monitoring equipment and sensor attachment methods were all selected and implemented in consultation with the MIT AgeLab and New England Univeristy Transportation Center to specifically support cross laboratory, country and cultural collaborative and replication studies (Son, Reimer, Mehler, Pohlmeyer, et al., 2010).

Figure 1. The DGIST driving simulator

Secondary task

An auditory delayed digit recall task was used to create periods of cognitive demand at three distinct levels. This form of n-back task requires participants to say out loud the nth stimulus back in a sequence that is presented via audio recording. (See Mehler et al. (2010) for additional background and information on different n-back task formats.) The lowest level n-back task is the 0-back where the participant is to immediately repeat out loud the last item presented. At the moderate level (1-back), the next-to-last stimuli is to be repeated. At the most difficult level (2-back), the second-to-the-last stimulus is to be repeated. The n-back was administered as a series of 30 second trials consisting of 10 single digit numbers (0-9) presented in a randomized order at an inter-stimulus interval of 2.1 seconds. Each task period consisted of a set of four trials at a defined level of difficulty resulting in demand periods that were each two minutes long.

Procedure

Following informed consent, sensor attachment and completion of a pre-experimental questionnaire, participants received 10 minutes of driving experience and adaptation time in the simulator. The simulation was then stopped and participants were trained in the n-back task while remaining seated in the vehicle. N-back training continued until participants met minimum performance criteria. Performance on the n-back was subsequently formally assessed at each of the three demand levels with 2 minute breaks between each level. When the simulation was resumed, participants drove in good weather through 37km of straight highway. Minutes 5 through 7 were used as a single task driving reference (baseline). Thirty seconds later, 18 seconds of instructions introduced the task (0, 1 or 2-back). Each n-back period was 2 minutes in duration (four 30 second trials). Two minute rest/recovery periods were provided before presenting instructions for the next task. Presentation order of the three levels of task difficulty was randomized across participants. A 2 minute interval starting 30 seconds after the last task was used as the post-task reference period (recovery).

Dependent variables

Error rates on the n-back were used to confirm the extent to which different conditions represented periods of higher cognitive workload. Error rate was calculated as the percentage of items that subjects responded to with an incorrect number or give no answer. Average forward velocity was selected as an indicator of compensatory behavior, since drivers have been observed to reduce their speed to manage increasing workload (Harms, 1991; Horberry et al., 2006; Son et al., 2010). Standard deviation of lateral position is another frequently used driving performance measure (Sayer, Devonshire & Flannagan, 2007). Heart rate and skin conductance level have been shown to be useful for quantifying changes in workload prior to the occurrence of driving performance degradation (Lenneman & Backs, 2009; Mehler et al. 2009; 2010).

Analysis method

The significance of age and workload on the dependent variables was assessed. Statistical comparisons were computed using a repeated measures general linear model (SPSS, ver. 17). A Greenhouse-Geisser correction was applied for models that violated the assumption of sphericity. Differences among significant main effects were assessed using pairwise t-tests with a least significant difference (LSD) adjustment for multiple comparisons.

RESULTS

Secondary Task Performance

Error rates on the 0-back, 1-back and 2-back tasks during the non-driving and driving conditions appear in Table 1. The overal higher error rates while driving are in line with the position that the demands of the primary driving task reduced the cognitive resources available to invest in the n-back. Under both non-driving conditions and in the dual-task conditon, error rates increased as the level of cognitive task difficulty increased. Considering the dual task period there was a main effect of difficulty level on errors ($F(1.4,39.3)=55.9$, $p<.001$), a main effect of age ($F(1,28)=30.8$, $p<.001$) reflecting the higher error rate in the older drivers, and a task difficulty by age

interaction (F(1.4,39.3)=30.0, p<.001) reflecting the dramatic rise in error rate in the older drivers relative to the younger group at the 2-back level.

Table 1. Secondary Task Error Scores

|	Non-driving			Dual Task		
Age	0-back	1-back	2-back	0-back	1-back	2-back
25-35	0% (0)	0.74% (2.87)	6.25% (8.35)	0% (0)	4.63% (7.47)	7.08% (6.83)
60-69	0% (0)	3% (4.86)	33.13% (20.86)	0.67% (1.14)	6.48% (6.70)	37.29% (17.46)

* Note: Means with standard deviations in parantheses

Physiological response

Means and standard deviations for heart rate and skin conductance level are listed in Table 2 and means are displayed graphically in Figure 3. Considering the sample as a whole, there was a significant effect of demand level on heart rate and skin conductance (F(2.2,62.6)=26.6, p<.001, F(3.3,91.7)=13.9, p<.001). Compared to the driving baseline period, heart rate during the 0-back task was 1.5 beats per minute (bpm) higher, 3.4 bpm higher during the 1-back, and 6.5 bpm higher during the 2-back. Post hoc comparisons show that the incrmental differences between each of these periods (baseline to 0-back, 0-back to 1-back, 1-back to 2-back) are statistically distinguishable (p=.002, p<.001, and p=.001 respectively). Skin conductance values during tasks compared to baseline were 0.5, 0.9, and 1.1 micromhos higher for the 0-back, 1-back, and 2-back periods respectively. Post hoc comparisons show that the incremental differences between baseline and 0-back and 0-back and 1-back are statistically distinguishable (p=.015 and p=.007). Mean skin conductance for the 2-back was higher than during the 1-back, but this difference was not statistically significant (p=.140). While means were consistently lower for the older drivers, and the absolute magnitude of change was somewhat less, there was no statistically significant effect of age on heart rate and skin conductance (F(1,28)=1.4, p=.251 and F(1,28)=0.5, p=.467).

Table 2. Summary of Physiological Response Measures During Simulated Driving

|	Heart Rate (beats/min)		Skin Conductance Level (micromhos)	
	Younger (25-35)	Older (60-69)	Younger (25-35)	Older (60-69)
Baseline	78.5 (15.4)	74.6 (11.1)	7.5 (3.5)	6.9 (3.1)
0-back	80.9 (15.1)	75.2 (10.1)	8.1 (3.8)	7.2 (3.2)
1-back	83.2 (15.1)	76.7 (10.1)	8.7 (3.9)	7.5 (3.2)
2-back	86.3 (15.9)	79.8 (11.8)	8.9 (3.9)	7.6 (3.4)
Recovery	77.9 (14.4)	73.2 (9.1)	7.9 (3.9)	7.1 (2.8)

* Note: mean (standard deviation)
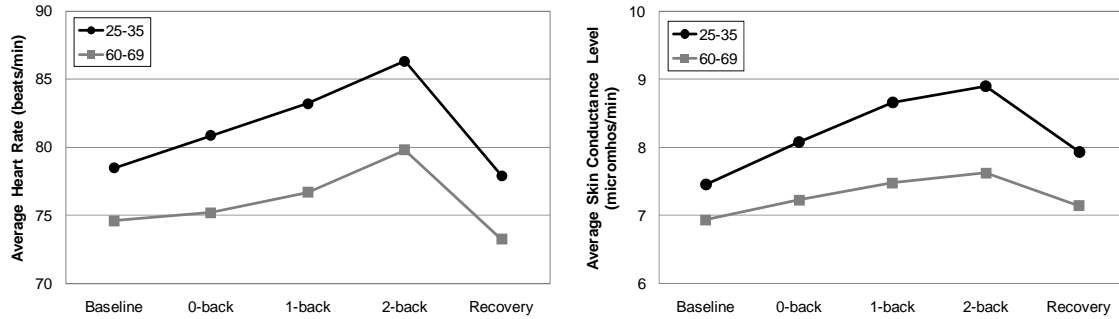
**Figure 3. Mean Physiological Arousal as a Function of Task Level:
Heart Rate (left) and Skin Conductance Level (right)**

Vehicle Control

To observe the compensatory behaviors and perfomance changes under different levels of cognitive workload by age, forward velocity and standard deviation of lane position were examined. As shown in Figure 4, both age groups decreased vehicle speed under the dual task conditions. Although the average velocity and the standard deviation of lane position profiles did not show a simple correlation with the level of cognitive workload, the driving performance was impacted during the dual task conditions as show in an overall effect of period (velocity: $F(3.1, 85.9)=13.5$, $p<.001$; SDLP: $F(3.4, 96.1)=6.3$, $p<.001$).
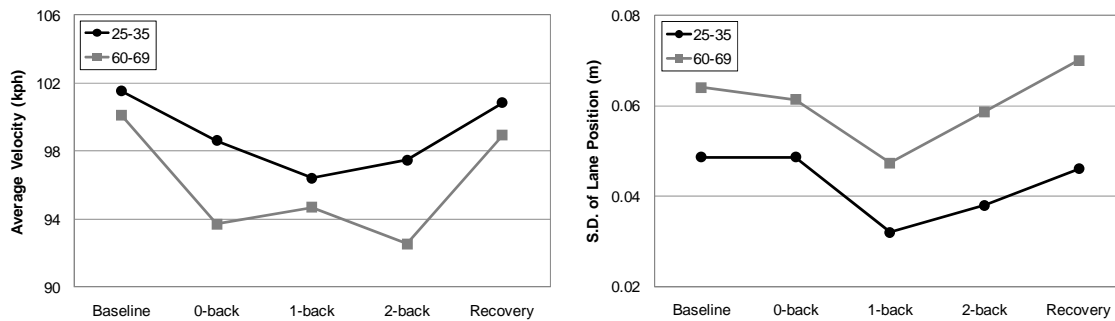
**Figure 4. Driving Performance Measures as a Function of Task Level:
Average Velocity (left) and Standard Deviation (SD) of Lane Position (right)**

DISCUSSION

N-back errors rates from both the non-driving period and while driving the simulator clearly document an incremental increase in difficulty across the task levels. Coincident with this, heart rate and skin conductance during simulated driving show an unambiguous increase in mean value for each level of heightened demand. These findings closely parallel what has been observed under field conditions (Mehler et al., 2010) and further strengthen the position that simulation can be used to validly model physiological reactivity patterns in response to incremental differences in cognitive workload (Reimer & Mehler, 2010). In addition, this study extends the available data on age and physiological activation while driving. As was the case in Mehler et al. (2010), individual variability in heart rate and skin conductance level are large enough (see standard deviation values in Table 2) that chronological age is not a statistically

significant factor in relatively healthy individuals up through the age of 69 who have been self-report screened for major medical complications that might affect reactivity. It remains to be determined to what extent the screened medications and health conditions would impact this finding. The pattern of change in heart rate and skin conductance level with increasing mental workload differs in important ways from what is seen in the two driving performance measures. Compensatory slowing of driving speed is observed during the dual task periods, yet the relationship to level of demand is not linear and could not be used, for example, to discriminate between the 0-back and 2-back tasks. Similarly, the impact of cognitive workload on standard deviation of lane position could not be used to correctly rank order task demand. These results support the arguments by Lenneman and Backs (2010) and Mehler et al. (2009) that selected physiological measures are more useful for detecting relative changes in mental workload at levels below which clear decrements in driving performance begin to be observed.

ACKNOWLEDGMENTS

REFERENCES

Brookhuis, K. A. & De Waard, D. (2001). Assessment of drivers' workload: Performance and subjective and physiological indexes. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, Workload, and Fatigue* (pp. 321-333). Mahwah, NJ: Lawrence Erlbaum Associates.

Brookhuis, K. A., de Vries, G. & de Waard, D. (1991). The effects of mobile telephoning on driving performance. *Accident Analysis and Prevention, 23,* 309-316.

Engström, J., Johansson, E., & Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F, 8(2),* 97-120.

Harms, L. (1991). Variation in drivers' cognitive load: Effects of driving through village areas and rural junctions. *Ergonomics*, 34(2), 151-160.

Horberry, T., Anderson, J., Regan, M. A., Triggs, T.J., & Brown, J. (2006). Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. *Accident Analysis and Prevention*, 38(1), 185-191.

Lenneman, J.K. & Backs, R.W. (2009). Cardiac autonomic control during simulated driving with a concurrent verbal working memory task. *Human Factors, 53*(3), 404-418.

Lenneman, J.K. & Backs, R.W. (2010). Enhancing assessment of in-vehicle technology attention demands with cardiac measures. *Proceedings of the Second International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutoUI 2010),* November 11-12, Pittsburgh, Pennsylvania, USA.

Mehler, B., Reimer, B., Coughlin, J.F., & Dusek, J.A. (2009). The impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record, 2138, 6-12*.

Mehler, B., Reimer, B., & Coughlin, J.F. (2010). Physiological reactivity to graded levels of cognitive workload across three age groups: An on-road evaluation. *Proceedings of the*

Human Factors and Ergonomics Society 54th Annual Meeting, San Francisco, CA, 2062-2066.

Reimer, B. & Mehler, B. (2010). The impact of cognitive workload on physiological arousal in young adult drivers: a field study and simulation validation. Manuscript under review.

Reimer, B., Mehler, B., Coughlin, J. F., Godfrey, K. M., & Tan, C. (2009). An on-road assessment of the impact of cognitive workload on physiological arousal in young adult drivers. *Proceedings of the AutomotiveUI*, Essen, Germany.

Reimer, B., Mehler, B., Wang, Y., & Coughlin, J.F. (2010). The impact of systematic variation of cognitive demand on drivers' visual attention across multiple age groups. *Proceedings of the 54th Annual Meeting of the Human Factors and Ergonomics Society*, San Francisco, Sept. 27-Oct. 1, 2010, 2052-2056.

Sayer, J., Devonshire, J. & Flannagan, C. (2007). Naturalistic driving performance during secondary tasks. *Proceedings of the Fourth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, 224-230.

Son, J., Reimer, B., Mehler, B., Pohlmeyer, A. E., Godfrey, K.M., Orszulak, J., Long, J., Kim, M.H., Lee, Y.T. & Coughlin, J.F. (2010). Age and cross-cultural comparison of drivers' cognitive workload and performance in simulated urban driving. *International Journal of Automotive Technology*, 11(4), 533-539.

Wilson, G. F. (2002). Psychophysiological test methods and procedures. In S. G. Charlton & T. G. O'Brien (Eds.), *Handbook of Human Factors Testing and Evaluation* (pp. 127-156). Mahwah, NJ: Lawrence Erlbaum.