THE TACTILE DETECTION RESPONSE TASK: PRELIMINARY VALIDATION FOR MEASURING THE ATTENTIONAL EFFECTS OF COGNITIVE LOAD

Richard A. Young, Li Hsieh & Sean Seaman
Wayne State University
Detroit, Michigan, USA
Email: ryoun@med.wayne.edu

Summary: Improved measures of the attentional effects of cognitive load are needed to reduce potential crashes caused by secondary tasks performed while driving. The Tactile Detection Response Task (TDRT) in the proposed ISO Draft Standard WD17488 was tested in laboratory and on-road venues with 16 and 15 subjects, respectively. A sensitivity test used a purely cognitive load increase from an easy (0-back) to hard (1-back) auditory-vocal task. The TDRT response time increased by 90±21 msec in the laboratory, and by 135±34 msec on the road, while the miss rate increased by 4% in the laboratory and 5% on the road, thus validating TDRT sensitivity to an increase in purely cognitive load. A specificity test used a visual load increase with little cognitive load difference from an easy to hard visual-manual "Surrogate Reference Task" (SuRT), to which the TDRT should not respond. The TDRT response time and miss rate to the SuRT did not increase in the laboratory or road as a result of the increased visual load, providing preliminary validation that the TDRT may be both specific and sensitive to the attentional effects of cognitive load.

INTRODUCTION

The U.S. National Highway Traffic Safety Administration has made driver distraction mitigation a major initiative, proposing guidelines for visual-manual tasks in vehicle-embedded systems in 2012,[1] and announcing plans for voice-based task guidelines in 2014. This classic dichotomy between visual-manual and auditory-vocal-cognitive tasks obscures the fact that visual-manual tasks, even highly-practiced ones,[2] also cause cognitive load, sometimes substantial.[3,4] Cognitive load likely increases crash risk by its effects on attentional networks in the brain that increase event response times (RTs), miss rates, and the duration of single glances off the road.[3,4,5]

The "Peripheral Detection Task" (PDT) has been proposed as one measure of the attentional effects of cognitive load.[1,3,6,7,8] The PDT measures RTs to, and misses of, peripheral light stimuli during secondary task performance. For auditory-vocal tasks not requiring eye movements (e.g. cell phone conversation), the PDT is sensitive to cognitive load effects on "orienting attention".[9] For visual-manual tasks requiring eye movements, the PDT is also sensitive to effects on attention networks from visual-perceptual load and manual-motor response conflicts.[3,4] However, the PDT is not *specific* to these attentional effects, because event detection also fails when the eyes move such that the PDT lights do not fall on the retina (a non-attentional effect).

Test methods and metrics that are both specific and sensitive to the attentional effects of cognitive load are thus needed for both visual-manual and auditory-vocal tasks. Preliminary results suggest that the Tactile Detection Response Task (TDRT), as specified by ISO Draft Standard WD17488,[10] may be one such test. Previous studies have demonstrated TDRT sensitivity to the attentional effects of cognitive load,[7,11,12,13,] but did not explicitly test for, nor interpret their results in terms of, specificity to those attentional effects. The current *objective* is to validate the TDRT as specific as well as sensitive for the attentional effects of cognitive load. The *strategy* is to test TDRT validity by using two task types identified by previous research: the first causes primarily cognitive load, while the second causes primarily visual load. Comparisons are then made between easy and hard versions of each task type to determine whether the TDRT is sensitive to cognitive load differences (*sensitivity test*[14]), and insensitive to visual load differences (*specificity test*[14]).

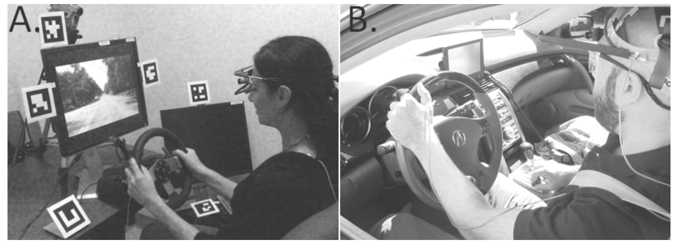

Figure 1. A. Laboratory venue. B. Road venue

METHOD

Primary Task and TDRT

Tests were conducted in the laboratory by Wayne State University (Fig. 1A), and on an open road (Fig. 1B) by Dynamic Research, Inc., using identical TDRT stimuli and methods as specified in ISO Draft Standard WD17488.[10] In the laboratory test, the primary task was for subjects to watch a high-definition video of real-world driving recorded from a driver's viewpoint and use the steering wheel to keep a cursor in the center of the lane.[15] At both sites, when subjects felt a "tactor" vibrate on the back of their left shoulder they pressed a microswitch on their left index finger against the steering wheel, which recorded the RT from stimulus onset. One tactile stimulus was presented at a time, with an inter-stimulus interval of 3 to 5 secs. The tactor turned off when the microswitch was pressed or after 1 sec, whichever came first. *Misses* were defined as RTs > 2500 ms, and were not included in the mean RT. RTs < 100 ms were discarded as invalid responses. The tactile stimuli and responses were controlled and recorded by a TDRT device purchased from TNO.[16]

Secondary Tasks: Descriptions, Resource Analyses, and Validation Criteria

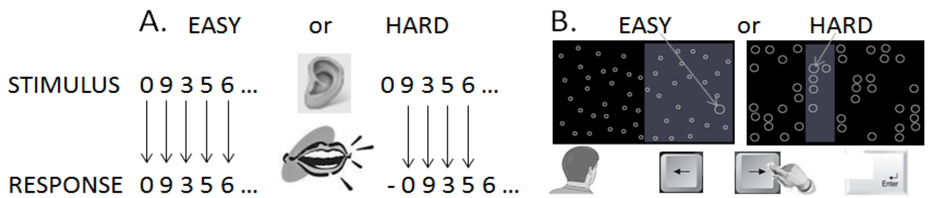

Figure 2. Secondary tasks. A. Auditory-vocal (n-back). B. Visual-manual (SuRT)

N-Back Task. Fig. 2A illustrates the n-back verbal digit memory recall task. It has a fixed digit presentation interval designed to measure continuous cognitive performance in a given time period. Subjects were presented a series of spoken digits for a 1.5 minute period on the road and a 2 minute period in the laboratory. In the 0-back (easy) task, they verbally repeated the last number heard. In the 1-back (hard) task, they repeated the number that was heard just before the last number heard.

Resource analysis: 1-back is known from prior research[17,18] to cause a purely cognitive load increase compared to 0-back. Both difficulty levels require storing and retrieving digits from verbal working memory, which involves the executive attention network.[9] The only resource difference is a slightly increased cognitive load on verbal working memory[17] from remembering two digits in the 1-back task rather than one digit in the 0-back task. The n-back task has no visual or manual load. The auditory (listening) and verbal (vocalizing single digits) loads are identical for 0-back and 1-back, so any effect of those loads on the TDRT metrics is removed analytically by subtracting the mean RT and %miss of 0-back from 1-back for each period for each subject. Sensitivity validation criterion: The TDRT is sensitive if 1-back shows increased RT and %miss to tactile events vs. the same measures for 0-back (paired t, 2-tailed $p < 0.05$).

Surrogate Reference Task (SuRT). Fig. 2B shows the SuRT visual search and response task.[19,20] On a screen to their right (Fig. 1), subjects scanned a display of circles with one larger "target" circle with a bolder stroke-width than the "distractor" circles (gray arrows, Fig. 2B). The target and distractor circle differences were easier to discriminate for easy vs. hard SuRT. Subjects pressed the left-right keypad buttons to move the gray outline bar to the region which contained the target circle and pressed the "enter" key to confirm their selection. The easy SuRT (Fig. 2B left) had 2 sections and the hard SuRT (Fig. 2B right) 6 sections to choose from. Subjects were instructed to repeat the task continuously during the test period and to complete each task repetition as quickly and accurately as possible, self-paced. Resource analysis: SuRT has high visual load with no auditory or verbal load. SuRT is thought to have little cognitive load,[20, p. 106] but even with high cognitive load, as long as it was about the same for hard vs. easy SuRT it would cancel out in the analysis after taking the difference of hard vs. easy TDRT responses. Specificity validation criterion: The TDRT is specific if hard vs. easy SuRT has no statistically significant increase in RT or %miss to tactile events (paired t, 2-tailed $p > 0.05$).

Baseline. Comparisons to a dual-task baseline (i.e., driving + TDRT) were not of interest. A triple-task condition (driving + TDRT + n-back) compared to a dual-task baseline (without n-back), has added auditory and verbal load in addition to added cognitive load, muddying any attempt to test sensitivity to the attentional effects of a purely cognitive load increase. Likewise, a triple-task condition (driving + TDRT + SuRT) compared to a dual-task baseline (without SuRT), has added cognitive load from task-switching costs, muddying any attempt to test TDRT specificity (i.e., insensitivity to visual load differences).

Subjects, Metrics and Procedures

Subjects. Healthy subjects 25-45 years old, each of whom drove >6,200 miles annually, were selected (8 males and 8 females for the laboratory test; 8 males and 7 females for the road test).

Metrics. RT and %miss for the tactile stimulus were used to validate sensitivity and specificity to the attentional effects of cognitive load. Secondary task performance metrics (error rate for n-back and SuRT, and task time and button presses for individual SuRT trials), and subjective workload estimates, were also tabulated. Driving performance metrics (variation of cursor in lane in laboratory; steering, yaw rate, and speed on road), and eye glance data were also collected.

Procedures. Procedures at laboratory and road sites were the same. Subjects filled out required informed consent paperwork, and TDRT training was conducted until subjects reported feeling comfortable with it. Training on the primary driving task with TDRT was conducted for at least 10 minutes. Instructions were: "Your main priority is to drive safely. Please remember to maintain your position in the center of the lane. The tactile and the [n-back/SuRT] tasks will both be active during the run. Please do your best to pay attention to both tasks." Then: (1) The first secondary task was reviewed and practiced for at least 30 seconds or 2-5 trials; (2) the data collection period began; (3) driving alone and secondary tasks while driving were performed for a 2-minute period in the lab and for a 1.5-minute period on road; (4) subjective workload and situation awareness self-ratings were collected using a simplified NASA TL/X scale[3]; (5) steps 1-4 were repeated for another data collection period; (6) steps 1-5 were repeated for each secondary task. The 4 secondary tasks were presented in randomized order for each subject. A systematic data analysis procedure was followed: (1) All RTs in a period were averaged to give a mean RT per period; (2) the %miss for each period was calculated by taking the total misses divided by total events; (3) the 2 repeated periods were averaged to give a mean RT and %miss per subject, per task; (4) means across subjects for RT and %miss for each task were calculated and plotted; (5) the difference score of each subject between the hard vs. easy tasks was tabulated, and the mean and standard error across those difference scores were calculated; (6) 2-tailed paired *t*-tests[21] were conducted for planned comparisons of 1-back minus 0-back differences (sensitivity test), and hard minus easy SuRT differences (specificity test).

RESULTS

Fig. 3 (left) shows that RT increased for 1-back (gray bars) vs. 0-back (white bars) (lab Δ 90±21 msec; road Δ 135±34 msec), validating TDRT sensitivity to the attentional effects of a purely cognitive load increase. Fig. 3 (right) shows that RT did not increase for the more visually-demanding hard SuRT (gray bars) vs. easy SuRT (white bars) (lab Δ -3±201 msec; road Δ -27±28 msec), thus validating TDRT specificity. The pattern of results for miss rate (percentages at bottom of each bar) was consistent. Error bars are ±standard error of subject paired RT differences. Of interest is the consistency for lab and road of the pattern of changes in RT and %miss across task difficulty levels; the absolute values of metrics between lab and road rarely agree,[3,8,22] and are of little interest here.

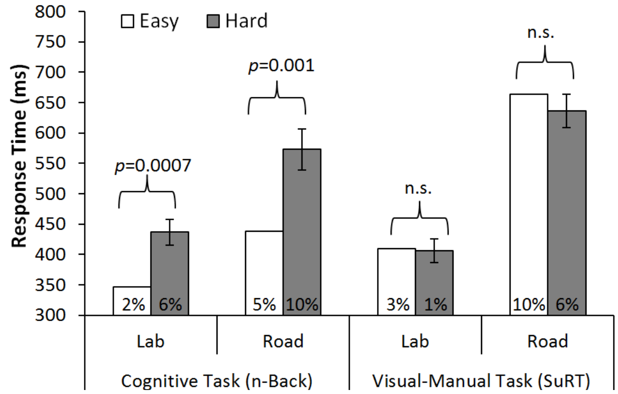

Figure 3. Responses to TDRT during secondary task performance in lab and on road. Left: Cognitive task. Right: Visual-manual task

Other measures of primary and secondary task performance indicated subjects were not differentially "shedding" any task. The laboratory data showed no statistically significant differences in lane-tracking performance among conditions. All tasks were completed with almost no errors in n-back, although the self-rated composite subjective workload score was about 3 times higher for 1-back vs. 0-back, supporting its greater difficulty. The added visual search difficulty for hard vs. easy SuRT was supported by 6 metrics: (1) 2 times higher subjective workload ratings; (2) 25% lower self-rated situation awareness; (3) 3 times slower mean task time per trial, 5.4 vs. 1.6 seconds; (4) 4 times more errors, 4.5% vs. 1.1%; (5) fewer trials completed in a 2-minute period, 24 vs. 87; and (6) a tendency for more eyes-off-road time. Conversely, there was increased manual load for the easy SuRT, with 40% more buttons pressed in a 2-minute period (94 for easy vs. 67 for hard SuRT). This pattern of primary and secondary task performance metrics was similar for the road data (not shown).

DISCUSSION AND CONCLUSION

The foregoing results help validate the sensitivity and specificity of the TDRT to the attentional effects of a purely cognitive load increase within the limited set of tasks examined in this research. Sensitivity was shown for a hard vs. easy cognitive task (auditory, vocal, visual, and manual loads held constant). Specificity was shown by a lack of sensitivity to a visual load increase for a hard vs. easy visual-manual task (auditory and vocal loads held constant). These results are consistent with a simulator study using similar tasks,[11] but the simulator study[11] did not explicitly test for specificity, nor interpret its results in terms of it. At least 8 explanatory hypotheses for this newly-recognized TDRT specificity effect are possible.

Competing Alternative Explanatory Hypotheses for the TDRT Specificity Effect

1. *Pop-out.* Unusually fast and easy detection of the larger SuRT circle may have occurred because it differed only in size and stroke width, permitting *pop-out*, or rapid perception of a unique feature in a large field of distractor features, which tends to occur in feature-based visual search tasks.[23]

Hence, there may have been little actual difference in visual perceptual load between hard and easy SuRT. This hypothesis is rejected because it has been shown that "ellipses varying in size in steps that were subjectively approximately equal,"[23, Fig. 4] did not pop-out, and required "serial fixations with fovea1 vision."[23, p. 113] The long mean search time for the hard SuRT target circle (5.4 sec) is also far longer than the 3 sec upper limit for pop-out.[23, p. 112]

2. *Classic dichotomy.* Auditory-vocal tasks cause cognitive load, but visual-manual tasks do not.[1] This hypothesis is rejected because the PDT is sensitive to visual-manual tasks with equal visual-manual loads but differing cognitive loads. An"HVAC" task with equal visual-manual load but larger cognitive load than a "Cassette" task had worse PDT scores.[4, Fig. 6; 8]
3. *Low sensitivity.* The hypothesis that TDRT lacks sensitivity to small differences between hard and easy SuRT is rejected because the TDRT was sensitive to the attentional effects of the known small cognitive load difference between 0- and 1-back tasks.[17,18] Hence, the null result for SuRT was not because of low sensitivity.
4. *No difference in visual-motor load.* Hard SuRT has more visual load, but easy SuRT has more motor load because of more button presses and eye glances.[11, Fig. 8] Hence, the sum of the visual and manual loads for easy SuRT may equal that for hard SuRT. However, 6 metrics indicated poorer secondary task performance for hard vs. easy SuRT (see Results).
5. *Lavie Load Model.*[24] This model states, "The ability to focus attention improves under task conditions of high perceptual load, but deteriorates under conditions of high load on cognitive control processes such as working memory."[24] It correctly predicts: (1) hard SuRT with higher perceptual load improves or at least does not deteriorate TDRT performance vs. easy SuRT; (2) 1-back with higher load on working memory deteriorates TDRT performance more than does 0-back.[25] However, the Lavie Load Model does not readily explain why both SuRT levels had worse RT performance than both n-back levels on the road (Fig. 3).
6. *Response interference*. RT may have increased for easy vs. hard SuRT because the greater number of right-hand button presses for the easy SuRT caused more response conflict with the left-hand button responses for the TDRT and the arm responses for steering, increasing executive attention load.[9] This would be particularly true on-road, where there is even more response conflict because of the increased steering and added foot responses for braking.[20, p. 106] Note also that there are known large individual differences and variability in susceptibility to response interference.[26] For example, a preliminary report[27] on the first 12 lab subjects and first 5 road subjects in this study found statistically significant TDRT RT decreases for hard vs. easy SuRT that disappeared (Fig. 3) with 4 more lab subjects and 10 more road subjects.
7. *Task management strategies by executive attention*. Because SuRT was self-paced (and n-back machine-paced[18]), subjects may have tried to strategically manage their workload in a way that was intended to preserve their overall performance on the driving and TDRT tasks, at the expense of hard SuRT performance. This explanation is supported by the 3 times slower pace of the hard SuRT, to an extent that would allow drivers to accomplish the more demanding visual search at a rate that could be interleaved with driving and TDRT, without compromising their performance too much. Hence, subjects could maintain equal TDRT performance for the two levels of SuRT but not the two levels of the n-back task.
8. *Driver performance dimension model.* Principle component analysis of driver performance metrics from nearly 100 on-road visual-manual tasks[3,4] found that physical demand metrics (number of task steps, eyes-off-road time, lane deviations, task time, etc.) cluster along dimension 1 (D1). Mental demand metrics (RT, %miss, long single glances) cluster along an orthogonal dimension 2 (D2). Many tasks score highly on D1 but not D2 (e.g. destination entry[3,4]), and SuRT may be this type. Other tasks score highly on D2 but not D1,[3,4] and n-back may be this type. However, this hypothesis is valid only if it can be safely assumed that the dimensional model (developed with visual-manual tasks) extends to auditory-vocal tasks.

Hypotheses 1-4 were rejected based on current results. Future studies should allow a choice between hypotheses 5-8 by testing a wider variety of tasks with different types and amounts of perceptual, cognitive, and motor loads (e.g., two tasks with equal visual load but unequal cognitive load, with vocal instead of manual responses). Also, if the goal is to calibrate or validate the DRT, all tasks should be machine-paced to ensure distinct and objectively-defined levels of those loads. Once validated, the DRT can be extended to more ecologically valid self-paced in-vehicle tasks.

In conclusion, under the conditions and the limited set of tasks used in this experiment, the results provide a preliminary validation that the TDRT is both specific and sensitive to the attentional effects caused by differences in cognitive load.

ACKNOWLEDGMENTS

This research was supported in part by the Toyota Collaborative Safety Research Center. We thank D. Chiang and A. Kirsch of Dynamic Research, Inc. for their road data; L. Angell, K. Ebe, J. Engström, J. Harbluk, J. Foley, N. Lavie, B. Mehler, M. Posner, and D. Raveh for comments.

REFERENCES

[1] National Highway Traffic Safety Administration. (2012). *Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices.* http://www.gpo.gov/fdsys/pkg/FR-2012-02-24/pdf/2012-4017.pdf.

[2] Groeger, J. A., & Clegg, B. A. (1997). Automaticity and driving: time to change gear conceptually. In *Traffic and Transport Psychology: Theory and Application*, Pergamon.

[3] Young, R. A., & Angell, L. S. (2003). The dimensions of driver performance during secondary manual tasks. *Proceedings of the Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Park City, Utah.

[4] Young, R. A. (2012). Event detection: The second dimension of driver performance for visual-manual tasks. *SAE Int. J. Passeng. Cars - Electron. Electr. Syst., 5*(1).

[5] Young, R.A. (2012). Cognitive distraction while driving: A critical review of definitions and prevalence in crashes. *SAE Int. J. Passeng. Cars - Electron. Electr. Syst.*, 5(1).

[6] Olsson, S., & Burns, P. C. (2000). *Measuring driver visual distraction with a peripheral detection task*. NHTSA Distraction. Forum, http://www-nrd.nhtsa.dot.gov/departments/Human%20Factors/driver-distraction/PDF/6.PDF.

[7] Engström, J., Åberg, N., Johansson, E., & Hammarbäck, J. (2005). Comparison between visual and tactile signal detection tasks applied to the safety assessment of in-vehicle information systems. *Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Rockport, Maine.

[8] Angell, L., Auflick, J., Austria, P., Biever, W., Diptiman, T., Hogsett, J., Kiger, S., Kochhar, D., Tijerina, L. (2006). Driver Workload Metrics Project, Task 2 Final Report. NHTSA.

[9] Posner, M. I., & Fan, J. (2008). Attention as an organ system. In J. R. Pomerantz (Ed.), *Topics in Integrative Neuroscience: From Cells to Cognition* (pp. 31–61). Cambridge University Press.

[10] ISO WD17488. (2012). Detection-Response Task for assessing selective attention in driving. ISO/TC 22/SC 13/WG 08 (preliminary work item, version 8, February).

[11] Mattes, S., Föhl, U., & Schindhelm, R. (2008). Empirical comparison of methods for off-line workload measurement, AIDE deliverable D2.2.7. EU project IST-1-507674-IP.

[12] Engström, J. (2010). The Tactile Detection Task as a method for assessing drivers' cognitive load. In G. L. Rupp (Ed.), *Performance Methods for Assessing Driver Distraction: The Quest for Improved Road Safety* (pp. 90-103). Warrendale, PA: SAE International.

[13] Diels, C. (2011). Tactile Detection Task as a real time cognitive workload measure. In M. Anderson (Ed.), *Contemporary Ergonomics and Human Factors* (pp. 183-190): CRC Press.

[14] *Sensitivity* and *specificity* are usually assessed by analyzing test results for individual subjects, but are adapted here for assessment of event detection results (see http://en.wikipedia.org/wiki/Sensitivity_and_specificity).

[15] Hsieh, L, Young, R, & Seaman, S. (2012). Development of the Enhanced Peripheral Detection Task: A surrogate test for driver distraction. *SAE Int. J. Passeng. Cars - Electron. Electr. Syst*., 5(1). doi: 10.4271/2012-01-0965. Available at: http://papers.sae.org/2012-01-0965/

[16] For purchase of the TDRT, contact Ronald.vanbuuren@tno.nl at TNO (http://www.tno.nl/).

[17] Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping, 25*(1), 46-59.

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

[19] ISO PDTS14198. (2010). Calibration tasks for methods which assess demand due to the use of in-vehicle systems. ISO/TC 22/SC 13/WG 08 (preparatory STD Version 2.2).

[20] Östlund, J., Nilsson, L., Törnros, J., & Forsman, Å. (2006). *Effects of cognitive and visual load in real and simulated driving*. Linköping, Sweden: VTI.

[21] The paired-*t* test produces identical results to a repeated-measures ANOVA with 2 groups.

[22] Young, R. A., Aryal, B., Muresan, M., Ding, X., Oja, S. & Simpson, S. N. (2005). Road-to-lab: Validation of the static load test for predicting on-road driving performance while using advanced information and communication devices. *Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Rockport, Maine. http://drivingassessment.uiowa.edu/DA2005/PDF/35_DickYoungformat.pdf.

[23] Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97-136.

[24] Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current Directions in Psychological Science, 19*(3), 143-148. doi: 10.1177/0963721410370295.

[25] Angell, L. S. (2010). Conceptualizing effects of secondary task demands during driving: Surrogate methods and issues. In G. L. Rupp (Ed*.), Performance Metrics for Assessing Driver Distraction: The Quest for Improved Road Safety* (pp. 42-72). Warrendale, PA, USA: SAE International.

[26] Forstmann, B., van den Wildenberg, W., & Ridderinkhof, K. (2008). Neural mechanisms, temporal dynamics, and individual differences in interference control. *Journal of Cognitive Neuroscience, 20*(10), 1854-1865.

[27] Young, R., Seaman, S., & Hsieh, L. (2012). *Measuring cognitive distraction on the road and in the lab with the Wayne State Detection Response Task.* Transportation Research Board 2012 Annual Meeting, Washington, D.C. Available at: http://amonline.trb.org/211ign/211ign/1.