MIT Project on Assessing the Demands of Voice Based In-Vehicle Interfaces: Study 2

White Paper 2014-7

An Executive Summary:

Further Evaluation of the Effects of a Production Level "Voice-Command" Interface on Driver Behavior: Replication and a Consideration of the Significance of Training Method

July 9, 2014

Bruce Mehler and Bryan Reimer

This document provides an overview / executive summary of the complete technical report:

Mehler, B., Reimer, B., Dobres, J., McAnulty, H., Mehler, A., & Coughlin, J. F. (2014). Further Evaluation of the Effects of a Production Level "Voice-Command" Interface on Driver Behavior: Replication and a Consideration of the Significance of Training Method. (Technical Report 2014-2). Cambridge, MA: MIT AgeLab.

Lead Project Contact:

Bryan Reimer, Ph.D.
Phone (617) 452-2177
reimer@mit.edu

Massachusetts Institute of Technology > AgeLab > 77 Massachusetts Ave, E40-291, Cambridge, MA 02139
Phone: 617.452.2177 > fax: 617.258.7570 > reimer@mit.edu > agelab.mit.edu

Executive Summary for Full Technical Report

In Study 1 of this project (Reimer, Mehler, Dobres & Coughlin, 2013), a sample of 60 participants equally balanced by gender and across two age groups (20-29 and 60-69 years) was evaluated in terms of self-reported workload, physiological arousal, visual attention, and driving performance metrics while engaging in a wide-range of voice-command based driver-vehicle interface (DVI) tasks in a commercially available production system under actual highway driving conditions. Among other findings, the data collected suggested that voice recognition was relatively robust, the cognitive load associated with the voice-command DVI's assessed was less than anticipated, and that use of the voice-based DVI for radio tuning resulted in less visual demand than the more traditional visual-manual interface. On the other hand, some "voice-command" interactions, particularly full destination address entry into the navigation system, proved to be highly multi-modal in nature and resulted in significant visual engagement when assessed in terms of total off-road glance time. If either the current National Highway Transportation Safety Administration (NHTSA, 2013) visual-manual distraction guidelines or the Alliance of Automotive Manufacturers' (Driver Focus-Telematics Working Group, 2006) guidelines for driver interactions with advanced in-vehicle information and communication systems (criterion 2.1A) were applied to the task, the DVI under study may not have met either guideline. In this regard, it is important to keep in mind that NHTSA's (2013) statement on the guidelines specifically indicated that they "are currently not applicable to the auditory-vocal portions of human-machine interfaces of electronic device", the system tested was put into production prior to the release of the NHTSA guidelines, and that the evaluation was done on-road and not in a simulator and in the exact form specified by NHTSA. Visual engagement values might be expected to be somewhat different following the NHTSA's protocol.

The present study (Study 2) assesses the extent to which key findings from Study 1 replicate in a second sample, as well as considering whether two differing approaches to introducing drivers to use of the DVI impact their pattern of interaction, including driving behavior. An analysis sample of 64 participants, equally balance by gender and distributed across the four age groupings (18-24, 25-39, 40-54, and 55+) recently specified by NHTSA (2013) for DVI assessment, was again evaluated during traditional manual radio tuning, voice-command assisted radio tuning, and voice-command assisted navigation system interaction consisting of full destination address entry and route cancelation. Participant behavior when presented with three levels of MIT

AgeLab auditory presentation / vocal response n-back cognitive demand reference task (0-, 1-, and 2-back) was also assessed. A new form of the n-back task, the "blank-back", was also introduced to assess behavior associated with just "listening" to n-back task auditory stimuli without the demand to hold any information in memory or make a verbal response.

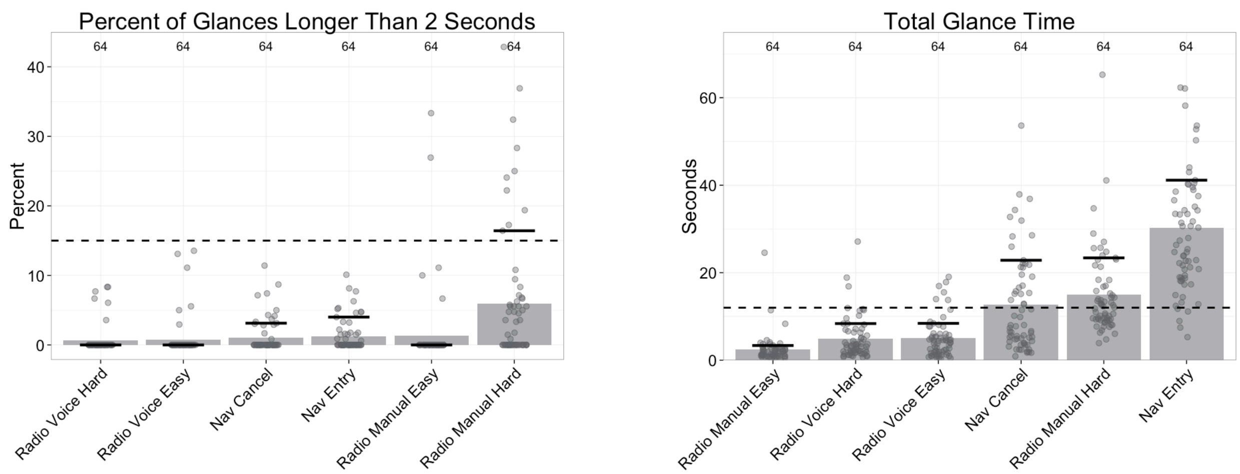
The two training methods evaluated consisted of essentially the same experimenter assisted, structured training protocol utilized in Study 1 and a self-directed experimentation period in which participants were informed of the DVI tasks they would be asked to consider attempting on-road and encouraged to explore the DVI to practice how to complete the tasks. During the self-guided training period, participants had access to the DVI user's guide and were taken through the voice-calibration procedure which developed familiarity with the press-to-talk voice interface button, but were given no other basic experimenter support in learning how to use the DVI. None of the participants in the self-guide training group was observed to pick-up and refer to the user's guide.

Contrary to expectations, no statistically significant main effects of training condition were observed across any of the primary outcome measures. Subjective impressions of the two experimenters who conducted the DVI training and on-road portions of the study suggest that the learning styles of individual participants varied greatly; it is possible that the distribution of participants that found each type of training beneficial or hindering was such that no net effects across training groups was observed. One consequence of the lack of statistically significant difference across training approaches was that the data collected in both training groups could be combined into a single large sample of 64 participants for purposes of assessing overall driver behavior in response to the various tasks under study.

A detailed analysis of individual tasks is presented in the body of report. The overarching finding is that the basic pattern of results seen in Study 1 (considering self-reported workload, physiological arousal, driving performance metrics, and glance metrics) largely replicate in the current study. Specifically:

- Voice recognition was found to be fairly robust with only 3 out of more than 80 participants unable to participate due to voice recognition issues.
- Apparent cognitive processing demand / workload as assessed through heart rate and skin conductance level (SCL) for the DVI tasks studied fell below the level of the 1-back cognitive reference task.

- For the radio tuning reference task (Radio Manual Hard), the voice-command method was associated with lower workload (self-report, heart rate, SCL), lower mean glance durations, a markedly lower percentage of long duration glances (Figure 1a), and significantly lower total glance time than the visual-manual interface (Figure 1b).
- Voice-command involved entry of a full destination address into the navigation system was associated with total eyes off-road time (TEORT) significantly above the acceptable criterion defined by NHTSA for visual-manual DVIs (Figure 1b) if the criterion was to be applied to this task.



a) b)

Figure 1: Representative plots showing glance time away from the forward roadway (with bars representing means and solid circles individual participants) across voice-command and manual radio tuning tasks as well as voice-command destination entry and route cancelation. If they were to be applied, the dashed lines indicate NHTSA visual-manual guideline criteria thresholds and the horizontal bars indicate where 85% of the participants fall relative to the threshold where: a) represents the 15% criterion for the percentage of single long duration (>2s) glances and b) the 12 second criterion for total glance time. (See full sized Figures 84 and 98 in the full report.)

As already noted, the NHTSA guidelines (NHTSA, 2013) for visual-manual distraction are explicit in stating that the guidelines "are currently not applicable to the auditory-vocal portions of human-machine interfaces of electronic devices." NHTSA has since released a supplementary report, *Explanatory Material About the Definition of a Task Used in NHTSA's Driver Distraction Guidelines, And Task Examples* (Angell, Perez & Garrott, 2013), that states "Some task interactions involve mixed-mode interactions: a mixture of both auditory-vocal and visual-manual interactions. Because such tasks do involve some visual-manual interaction, it is appropriate that the visual-manual components of these tasks meet the proposed Phase 1 NHTSA Distraction Guidelines" (p. 32). It is not entirely clear from these two statements whether the visual-manual and auditory-vocal components of the destination entry task considered here should somehow be partitioned or if the visual engagement over the entire interaction should be considered. It seems inherent in the second statement that glances to the display when it presents the driver with a visual listing of destination options should be included in the calculation of glance metrics. On the other hand, if the driver looks at the display screen to confirm if the system correctly understood a verbal command at another point in the destination entry task, should this be excluded since it might be considered as associated with an auditory-vocal portion of the task? An alternate approach is to assume that this glance away from the roadway was associated with the overall DVI interaction and should be counted, which was done in the EORT analyses in this report. Perhaps an even grayer area is the question of the inclusion of other glances, such as mirror inspections, that may occur multiple times during extended mixed-mode interactions. If visual engagement is considered over the full length of the task, should the broader temporal components of multi-modal tasks be considered in future guideline work? This question is developed further below.

The results presented in this report confirm the finding from Study 1 that the voice-command interface assessed was found to be highly multi-modal in nature – involving a mixture of auditory, vocal, visual, and manual components. Further, ongoing work by our group considering voice-command DVI's in other vehicles strongly indicates that this observation is far from unique to the DVI assessed here. As found in Study 1, the individual glance characteristics during the navigation entry task (i.e. mean single glance duration and percentage of glances longer than 2 seconds) were well within NHTSA guidelines for visual-manual distraction (if applied to these voice-command involved tasks). On the other hand, the total glance time away from the forward roadway over the course of full address entry was significantly longer than the 12 second threshold specified in the visual-manual distraction guidelines if they are applied to this task; mean TEORT for the sample was 30 seconds and 15% of participants had TEORT values greater than 40 seconds. However, it is also apparent that these glances were distributed over a much longer time period than those seen in the manual radio tuning task that was used as a reference in developing the guidelines. While manual radio tuning in the present study had a mean task completion time of 28 seconds, the destination entry task was distributed across a mean task interval of 113 seconds (four times longer)(SD 31 seconds). Thus, while any glances away from the forward roadway are of potential concern, as are tasks that draw on attentional resources of any type for extended periods of time, these data for the destination entry task at least raise the question as to whether TEORT guidance developed for classic visual-manual interactions in Phase I are appropriately extensible to longer multi-modal DVI interactions that are now becoming more common control options for the driver.

Summarizing the key observations across Study 1 and Study 2, it is apparent in the assessment of the voice-command interface in these studies that the attentional draws in modern DVI's can be highly multi-modal. Depending upon a given design, newer DVIs can involve various combinations of demand (visual, manual, auditory, vocal, haptic, cognitive, etc.). Simply including voice in a DVI does not preclude the need to consider possible visual engagement characteristics that may be present. Visual as well as other potential demand sources need to be included in the assessment of voice-command interfaces. Based on the measures collected (self-reported workload, peripheral physiological arousal, basic driving metrics, and comparative surrogate cognitive tasks), cognitive demand may be relatively moderate in appropriately designed voice-command interactions. Cognitive demand becomes more apparent when drivers have difficulty completing activities ("song-fail" task in Study 1 or have difficulty recalling command syntax such as "navigation cancel route" in Study 2).The latter finding in Study 2 suggests that ensuring a user has an established understanding of a system's syntax is necessary when evaluating the fundamental cognitive demand of an interaction. This may need to be considered independently from the cognitive resources invested by a driver who finds the interaction non-intuitive.

As a potential limitation, it should be noted that the measures of cognitive demand in the present studies were not exhaustive. Further, it is conceivable that other voice-command implementations might be associated with overtly higher levels of cognitive demand. For example, while some of the support displays provided as part of the destination entry task introduced visual demand, allowing the driver to glance at a list of street name options might well be less cognitively demanding than presenting an auditory list of numbered options that a driver has to hold in memory prior to making a selection. This again highlights the importance to taking all potential sources of demand and attempting to evaluate the net effect of their interaction in developing effective DVIs.

As noted, the gender and age distribution characteristics of the participants in this on-road study fully comply with NHTSA guidelines for DVI evaluation. In this regard, it may be useful to note that statistically significant main effects of gender appeared in males showing longer mean single glance durations and a higher percentage of long duration (> 2 second) glances. Age showed a statistically significant main effect on TEORT, with eyes-off-road time being higher in the older age groups. Similar trends were seen in mean single glance duration and percentage of long duration glances. Thus, not unexpectedly, age and gender factors can influence assessment measures. NHTSA's recommendation that evaluation samples be balanced by age and gender, and that the age distribution include a representative sample of older drivers, provides an opportunity to assess such differences.

Detailed reporting of descriptive statistics for each task type, including break-downs by age group, gender, and training condition, is provided for each outcome measure. Statistical analyses are provided for primary research questions and general comparisons across tasks. A range of graphic representations of the data, including plots showing the distribution of individual participant behavior, are provided to aid in visualizing the results. For visual metrics, assessment based on both the NHTSA eyes-off-the-forward-roadway (TEORT) and more traditional glance-to-device (GTD) measures are provided. In addition, several summary results sections are provided:

- *Summary Findings Comparing Training Conditions*
- *Summary Findings Comparing Manual & Voice-Based Radio "Hard" Tuning*
- *Summary Findings Comparing Voice Nav. Entry and Manual Radio Tuning*
- *Summary Findings Considering Age*
- *Summary Findings Considering Gender*
- *NHTSA Glance Metrics Summary Table*
- *Summary Comparison of Task Time and Glance Metrics*

Additional conclusions, a discussion of limitations, and a consideration of next steps are also provided in the full technical report.

Acknowledgements

Acknowledgement is extended to the Toyota Collaborative Safety Research Center (CSRC) which provided the primary funding for this project. In addition, we are particularly grateful for the valuable, constructive comments that James Foley and Kazutoshi Ebe of CSRC provided during the development of the study.

Supplemental / matching support was provided US DOT's Region I New England University Transportation Center at MIT. The vehicle under test was purchased through funding from Ford Motor Company for an earlier project assessing the Ford Active Park Assist™ feature (Reimer, Mehler, & Coughlin, 2010) and support for originally instrumenting the vehicle was supplied by The Santos Family Foundation. The current project was conducted without consultation or involvement of Ford Motor Company.

This work would not have been possible without the support of AgeLab staff and visiting scholars including: Erin McKissick, Enrique Abdon Garcia Perez, Adrian Rumpold, Thomas Manhardt, Brahmi Pugh, Martin Lavalliere and Brendan Drischler in the development and refinement of data analysis and extraction tools, and exhaustive reduction and coding of eye glance and other data.

The interpretive aspects of this report reflect the views of the authors, who are also responsible for the factualness and accuracy of the information presented herein.

References

Angell, L., Perez, M., & Garrott, W. R. (2013, November). Explanatory material about the definition of a task used in NHTSA's driver distraction guidelines, and task examples. (Report No. DOT HS 811 858). Washington, DC: National Highway Traffic Safety Administration.

Driver Focus-Telematics Working Group. (2006). Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communication Systems, Version 2.0: Alliance of Automotive Manufacturers.

National Highway Traffic Safety Administration. (2013). *(Issued Guidelines)* Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices (Docket No. NHTSA-2010-0053). Washington, DC: U.S. Department of Transportation National Highway Traffic Safety Administration (NHTSA).

Mehler, B., Reimer, B., Dobres, J., McAnulty, H., Mehler, A., & Coughlin, J. F. (2014). Further Evaluation of the Effects of a Production Level "Voice-Command" Interface on Driver Behavior: Replication and a Consideration of the Significance of Training Method. (Technical Report 2014-2). Cambridge, MA: MIT AgeLab.

Reimer, B., Mehler, B., & Coughlin, J. F. (2010). An evaluation of driver reactions to new vehicle parking assist technologies developed to reduce driver stress (MIT AgeLab White Paper). Cambridge, MA: Massachusetts Institute of Technology.

Reimer, B., Mehler, B., Dobres, J., & Coughlin, J. F. (2013). The Effects of a Production Level "Voice-Command" Interface on Driver Behavior: Reported Workload, Physiology, Visual Attention, and Driving Performance (Technical Report 2013-17A). Cambridge, MA: MIT AgeLab.

ABOUT THE AUTHORS

Bruce Mehler, M.A.

Bruce Mehler is a Research Scientist in the Massachusetts Institute of Technology AgeLab and the New England University Transportation Center, and is the former Director of Applications & Development at NeuroDyne Medical Corporation. He has an extensive background in the development and application of non-invasive physiological monitoring technologies and research interests in workload assessment, individual differences in response to cognitive demand and stress in applied environments, and in how individuals adapt to new technologies. Mr. Mehler is an author of numerous peer reviewed journal and conference papers in the biobehavioral and transportation literature. He continues to maintain an interest in health status and behavior from his early work in behavioral medicine. He received an MA in Psychology from Boston University and a BS degree from the University of Washington.

bmehler@mit.edu
(617) 253-3534
http://agelab.mit.edu/bruce-mehler

Bryan Reimer, Ph.D.

Bryan Reimer is a Research Engineer in the Massachusetts Institute of Technology AgeLab and the Associate Director of the New England University Transportation Center. His research seeks to develop new models and methodologies to measure and understand human behavior in dynamic environments utilizing physiological signals, visual behavior monitoring, and overall performance measures. Dr. Reimer leads a multidisciplinary team of researchers and students focused on understanding how drivers respond to the increasing complexity of the operating environment and on finding solutions to the next generation of human factors challenges associated with distracted driving, automation and other in-vehicle technologies. He directs work focused on how drivers across the lifespan are affected by in-vehicle interfaces, safety systems, portable technologies, different types and levels of cognitive load. Dr. Reimer is an author on over 70 peer reviewed journal and conference papers in transportation. Dr. Reimer is a graduate of the University of Rhode Island with a Ph.D. in Industrial and Manufacturing Engineering.

reimer@mit.edu
(617) 452-2177
http://web.mit.edu/reimer/www/

About the New England University Transportation Center & MIT Center for Transportation & Logistics

The New England University Transportation Center is a research, education and technology transfer program sponsored by the US Department of Transportation. Together the faculty, researchers and students sponsored by the New England Center conduct work in partnership with industry, state & local governments, foundations and other stakeholders to address the future transportation challenges of aging, new technologies and environmental change on the nation's transportation system. For more information about the New England University Transportation Center, visit utc.mit.edu. For more information about the US Department of Transportation's University Transportation Centers Program, please visit www.rita.dot.gov/utc/. The New England Center is based within MIT's Center for Transportation & Logistics, a world leader in supply chain management education and research. CTL has made significant contributions to transportation and supply chain logistics and helped numerous companies gain competitive advantage from its cutting edge research. For more information on CTL, visit ctl.mit.edu.

About the AgeLab

The Massachusetts Institute of Technology AgeLab conducts research in human behavior and technology to develop new ideas to improve the quality of life of older people. Based within MIT's Engineering Systems Division and Center for Transportation & Logistics, the AgeLab has assembled a multidisciplinary team of researchers, as well as government and industry partners, to develop innovations that will invent how we will live, work and play tomorrow. For more information about AgeLab, visit agelab.mit.edu