

Linking the Detection Response Task and the AttenD Algorithm through the Assessment of Human-Machine Interface Workload

Joonbum Lee

MIT AgeLab and New England University Transportation Center
77 Massachusetts Avenue, E40-215, Cambridge, MA 02139
Tel: 617 715 2011 Fax: 617 258 7570; Email: joonbum@mit.edu

Ben D. Sawyer

MIT AgeLab and New England University Transportation Center
77 Massachusetts Avenue, E40-279, Cambridge, MA 02139
Tel: 617 420 2237 Fax: 617 258 7570; Email: bsawyer@mit.edu

Bruce Mehler

MIT AgeLab and New England University Transportation Center
77 Massachusetts Avenue, E40-279, Cambridge, MA 02139
Tel: (617) 253 3534; Email: bmehler@mit.edu

Linda Angell

Touchstone Evaluations, Inc.
440 Burroughs Street, Detroit, MI 48202
Tel: (313) 881 6641; Email: langell@touchstoneevaluations.com

Bobbie D. Seppelt

Touchstone Evaluations, Inc.
440 Burroughs Street, Detroit, MI 48202
Tel: (319) 331 9626; Email: bseppelt@touchstoneevaluations.com

Sean Seaman

Touchstone Evaluations, Inc.
440 Burroughs Street, Detroit, MI 48202
Tel: (313) 523 1737; Email: sseaman@touchstoneevaluations.com

Lex Fridman

MIT AgeLab and New England University Transportation Center
77 Massachusetts Avenue, E40-215, Cambridge, MA 02139
Tel: 617 715 2011 Fax: 617 258 7570; Email: fridman@mit.edu

Bryan Reimer

MIT AgeLab and New England University Transportation Center
77 Massachusetts Avenue, E40-279, Cambridge, MA 02139
Tel: (617) 452 2177 Fax: (617) 258 7570; Email: reimer@mit.edu

Word count: 5,962 words text + 6 tables/figures x 250 words (each) = 7,462 words

Original Submission Date: 08/01/2016 Revision Date: 11/15/2016

ABSTRACT

Failures in drivers' attention allocation become evident when multi-tasking related demands leave vehicle operators unable to detect or respond appropriately to roadway threats or interfere adversely with their ability to appropriately control the vehicle. Robust methods for obtaining evidence and data about demands upon and decrements in the allocation of driver attention are needed as input for design, training, and policy. The detection response task (DRT) is one method (ISO 17488, 2016) that has been forwarded as a method for measuring the attentional effects of cognitive load. The AttenD algorithm is a method intended to measure driver distraction through real-time glance analysis, in which individual glances are converted into a scalar value using simple rules considering glance duration, frequency, and location. In the present work, a relationship between the two tools is explored. A previous multi-tasking driving simulation study, which used the DRT to differentiate the demands of a primary visual-manual human-machine interface (HMI) from alternative auditory-vocal involved multi-modal HMIs, was reanalyzed using AttenD, and the two analyses compared. Results support an association between DRT performance and AttenD algorithm output. Summary statistics produced from AttenD profiles differentiate between the demands of the HMIs considered with more power than analyses of DRT response time and miss rate. Among discussed implications is the possibility that AttenD taps some of the same attentional effects as the DRT. Future research paths, strategies for analysis of past and future datasets, and possible application in driver state-detection are also discussed.

Keywords: Driver distraction, Detection response task, Human-machine interface, Cognitive demand

INTRODUCTION

Operating a vehicle is a task with great variability in demand due to the dynamic, complex environment within which it takes place. When secondary in-vehicle tasks draw on the same resources required for driving, there is increased potential for impaired multi-tasking, and either driving or in-vehicle task performance, or both, may degrade (1). Drivers commonly attempt to compensate for increased demand from multi-tasking; for example, by either maintaining a greater following distance (2) or by driving more slowly (3). Despite these demand-mitigating strategies, unexpected spikes in roadway demands, or demands of the in-vehicle task itself, can lead to overload, dynamic instability of the driving task, and subsequent failure to detect and avoid hazards (4). For the individual, elevated demand levels may be inconsequential so long as overall load remains stable and sufficient resources are available. For the same reason, driving detriment due to occasional high demand spikes may be difficult for researchers to detect unless extreme and often environmentally-invalid manipulations are deployed (5). Even then, commonly used measures such as variability in lane position and longer psychological refractory periods can be subtle or changeable in pattern (2). As such, there is a need for better assessment methods that provide less ambiguity in evidence of driver overload risk and associated distraction.

In the pursuit of robust indicators of overloaded driving, a number of approaches have been taken. Peripheral detection tasks (PDTs) were succeeded by detection response tasks (DRTs) (6), which have emerged in recent years as a method for measuring the attentional effects of cognitive load as represented by ISO standard 17488 (7). In this approach, a stimulus is presented every few seconds, and the driver is asked to respond to it when it is detected (using a small response button attached to an index finger on their left hand). The stimulus may be either tactile (a vibration applied to the skin of the shoulder) (tactile DRT, or TDRT), or visual, in the form of a small red light on a head-mounted “antenna” (HDRT, for head-mounted DRT), or a remotely mounted red LED (remote visual DRT, or RDRT). The DRT stimuli appear randomly every 3 to 5 seconds (measured from one stimulus onset to the next), and are sustained for an exposure duration of 1 second. The DRT task is designed to be presented concurrently while the driver performs other, secondary tasks. Measurements of response time (RT) and accuracy (hit/miss rates) are its outputs. Changes in these measures across conditions are interpreted to arise primarily from the cognitive load effects of the task being tested. This is particularly well supported for tactile and head-mounted DRTs, which minimize the need for glancing to see the DRT stimuli. However, it should be noted that every DRT has specific input modality characteristics – and the effects of a secondary task on the response metrics will depend on the configuration of demands needed to perform that DRT – in combination with the other tasks underway. For instance, if visual orientation in the general direction of the forward roadway is required to detect a remote DRT stimulus, this will be more greatly affected by visual secondary tasks than will a tactile or head-mounted DRT). Thus, a DRT could be more broadly characterized as assessing the effects of “task load” on attention – with some forms of DRT tending to minimize effects of any visual processing (e.g., the tactile DRT) – and other forms of DRT tending to reflect both visual and cognitive aspects of a secondary task’s effects on the DRT response (*for instance*, 8). Signal detection analysis approaches can help to refine such analyses by disambiguating the accuracy measure (9).

The value of such data is multi-faceted, but one promising use is in the evaluation of the demands associated with in-vehicle user interfaces (10). Several lines of research suggest that auditory-vocal interactions, while increasing demand placed on the driver relative to just driving, result in lower self-reported workload and divert visual resources from the roadway less than

their visual-manual counterparts (11-14). However, the extent to which elevated cognitive demand is present in such interactions has been raised (15-17). One body of work (17-19) has used the DRT to quantify the relative “cognitive demand” of various experimental auditory-vocal interactions as well as real-world auditory-vocal involved interfaces from the “structural demands” of “just driving” (although in the absence of comparison to alternate visual-manual interfaces for accomplishing the same goals). While such work provides data on impairments in responsiveness to the DRT during device interaction, it is difficult to fully assess the degree to which the DRT is assessing cognitive demand in isolation or in combination with visual processes in these studies. Findings on drivers’ interaction strategies with auditory-vocal interfaces have strong implications for effective vehicle human-machine interface (HMI) design, driver training, and legislation and regulation. As such, deeper understanding of the role of the DRT as an assessment metric and the relationship between the measure and other alternatives is critically needed.

The DRT methods covered in ISO 17488 have been shown to be quite sensitive to increases in cognitive load using objectively defined levels of working memory demand (*for instance, 8*). Nonetheless, the emphasis on the use of the DRT to detect the effect of cognitive load on attention leaves open the question of how cognitive load and visual demand considerations interact, or might be considered together, to estimate net demand for purposes of HMI comparisons. Further, the introduction of the response task itself is the introduction of further multi-tasking, and so the very act of using the DRT (or PDT) to observe driver behavior may exert an influence on that behavior. For example, a recent study (20) observed that mental workload was rated higher in simulated driving with visual or tactile DRTs than without. This is concerning for reasons of environmental validity. Furthermore, the DRT can only probe effects intermittently – every few seconds during engagement in a secondary task. It cannot give an unaltered picture of how attention is allocated during the primary driving task, or even moment-to-moment assessment during the period of multitasking itself.

Kircher and Ahlström introduced the AttenD algorithm to detect distracted driving based on the allocation of visual resources (21-23). The AttenD algorithm has promise as an unobtrusive measure, and, unlike the DRT measurement method, does not impose demands of its own on the driver. Instead, the AttenD metric is derived from measurements of the natural, unaltered glance patterns that take place while driving and when non-driving-related tasks are duly performed. In the AttenD model, when the calculated metric falls to zero, the driver is considered to be distracted, although thinking of this value as a binary threshold likely underrepresents the potential power of the measurement approach. As a distraction detection algorithm, AttenD has been validated with empirical data, and also compared with other vision-based distraction detection algorithms (24). Importantly, the AttenD algorithm provides a continuous measure over an epoch of driving or multi-tasking.

It should be made clear that Kircher and Ahlström (21-23) explicitly presented the AttenD algorithm as a method to provide a continuous indication of the “extent of inattention or distraction” shown by a driver. While not the primary focus of this paper, the authors have been involved in efforts that build on elements of AttenD, but that reconceptualize the resulting metric as a continuous measure of driving relevant attention across time, as opposed to the important but potentially more narrow concept of distraction. Both the aforementioned work and the original AttenD algorithm offer a potentially more complete picture of attention than does the intermittent detection of a DRT stimulus probe across such an epoch. Such moment-to-moment measures of driving relevant attentional resource allocation, which increase during glances to the road and decrease while the driver looks away, can additionally provide an index of driver

awareness of the road environment. Consequently, such an approach may be considered a potentially potent tool in understanding the impact of in-vehicle HMI on driver attention.

A key research question the present work addresses is: “If various HMIs used in-vehicle present a variable mix of visual, manual, auditory, vocal, cognitive, and perhaps other (e.g., haptic) demands, how might various interfaces most pragmatically be compared for their impact on overall driving relevant attention?”. HMIs available in the automobile are increasingly leveraging modalities beyond vision to present information to the driver (e.g., auditory), and control and selection modalities now extend beyond manual manipulation to frequently include voice-command options. It has further been argued that the traditional division of demand into siloes of visual, manual, and cognitive may not be the most useful approach given the reality of modern in-vehicle HMIs, noting that cognitive resources can be argued to be involved in all HMI interactions and that production interfaces that are often characterized as “cognitive” (e.g., auditory-vocal) might better be considered as multi-modal as they often involve significant draw on visual resources (25-27). With these considerations in mind, two seemingly distinct assessment methodologies – the DRT, which has been forwarded as a measure of the effect of cognitive load on attention – and the AttenD, reconceptualized as a measure of visual attention allocation – are compared in how they rank the attentional demands of actual user interfaces.

An exploratory effort was undertaken to use the AttenD algorithm to re-analyze gaze data from a simulator study on driver multi-tasking that used the remote DRT technique. Broadly, we wanted to understand if a continuous measure of attention captured through the AttenD algorithm and discrete sampling afforded by the DRT would provide similar interpretations of relative demand of across several HMIs. In comparing the two assessment approaches, this datasource supported consideration of a classically defined visual-manual HMI interaction as compared to two variants of an auditory-vocal HMI for completing the same secondary task goal.

METHODS

This study is a secondary analysis of data from (28), which investigated drivers’ use of different interface modes for destination address entry while driving. While complete methodological details can be found in the initial paper, key details related to the present manuscript are summarized next.

Participants

Participants were recruited across two age groups (20-24 and over 55) from the greater Boston area using online and newspaper postings. Participants were required to meet a number of criteria: (a) a valid driver’s license for more than three years, (b) driving on average at least one time per week, (c) being in self-reported reasonable good health for their age and meeting a set of health exclusion criteria, (d) clearly understanding and speaking English, (e) no police reported accident in the past year, and (f) not actively using any medications causing drowsiness. Compensation of \$40 was provided for participation. Out of the 24 participants from the original study analysis set, two cases had video image issues that precluded coding of the glance data, resulting in 22 cases being available for the main analysis of the present study (mean age = 46.18, $SD = 21.51$, min = 20, max = 68, n of female participants = 9, and n of male participants = 13).

Apparatus

The study utilized a fixed based driving simulator in the MIT AgeLab: a full cab Volkswagen New Beetle with a front projection system providing a view of approximately 40 degrees (Figure

1). Graphical updates were generated using STISIM Drive version 2.08.02 (Systems Technology, Inc., Hawthorne, CA) based upon a driver's interaction with the steering wheel, brake, and accelerator. Instructions and audio tasks were pre-recorded and presented through the vehicle sound system. Correspondence between the demands of this simulator configuration and actual driving scenarios has been established through previous research (29, 30). The driving scenario consisted of a two lane rural road, without curves and a posted speed limit of 50 mph. A CogLens remote mounted DRT was implemented in accordance with the ISO Standard (7). A red LED was mounted on the windshield near the center of the participant's field-of-view and responses were recorded from a micro-switch placed on the participant's left index finger. Following the standard, the LED was activated every 3-5 seconds for a period of one second or until the participant responded using the finger-mounted switch.



FIGURE 1 MIT AgeLab driving simulator at time of study (left) and the location of the DRT light (right).

Destination Entry Device

Participants entered destination addresses in a Samsung Galaxy S4, model number SCH-1545 (released March 2013), which featured a 5" display with 1920 x 1080 resolution. The mobile network carrier was Verizon Wireless and the operating system was Android 4.3. The device was "free floating" (not mounted) and participants held it in their hand or rested it on the center console or other location at their discretion while performing tasks. Navigation tasks were carried-out using the Google Maps application, and participants were extensively trained on how to enter an address using the three different interaction modes (Table 1). One method required visual-manual touchscreen interaction and two were auditory-vocal-visual-manual (e.g., voice based commands). For all three methods, participants were instructed to first wake up the phone by pressing the large home button at the bottom of the screen. To enter voice recognition mode, participants would then double-tap the home button. For the touch interface, participants opened the Google Maps application, typed a specified address into the search bar, selected a car icon to show driving routes, selected a route and tapped "Start navigation." In the "Hands-Free" mode, participants enabled the voice recognition feature by double pressing the phone's home button. The phone then presented a seemingly random introduction, e.g., "Hello, I hope you are having a great and productive week. If you need any help say 'Hi Galaxy'." After saying "Hi Galaxy," the participant would then say "navigate to," followed by the street address. In the "standard" voice mode, the verbose audio introduction and "Hi Galaxy" command were omitted; after double-tapping the voice button, participants would immediately speak the navigation command.

AttenD Algorithm

Scores are generated from a 2-second time buffer, increasing while a driver looks on the road and decreasing while the driver looks away from the road. The closer the buffer falls to zero, the closer the driver is assumed to be distracted. More details of the algorithm can be found in Kircher and Ahlstrom's original manuscripts (21-23), which we used to perform buffer calculations. Figure 2 illustrates one participant's buffer profile during the manual destination entry task.

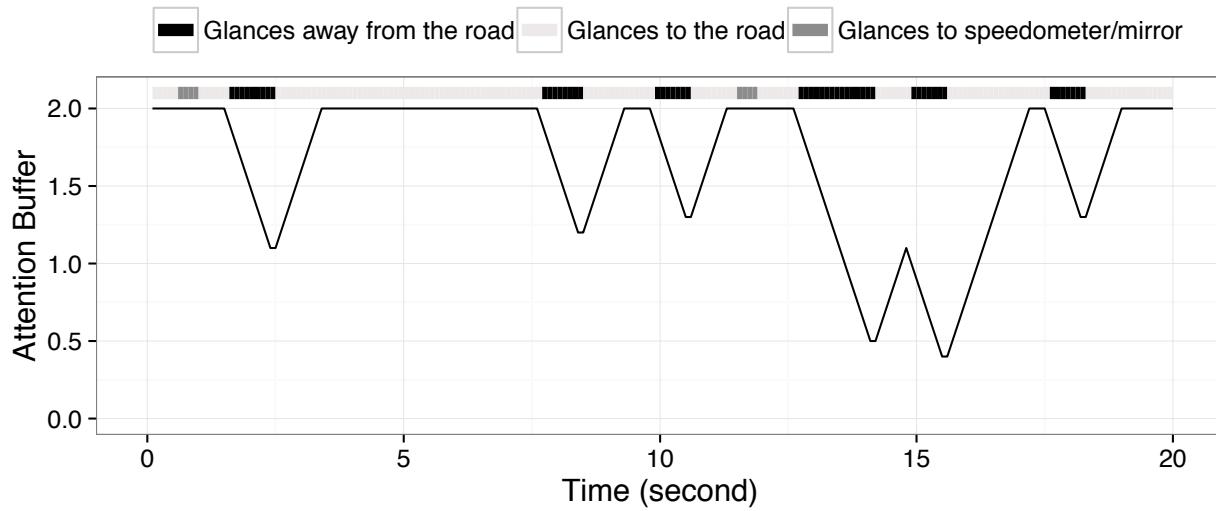


FIGURE 2 Example of a buffer profile over time (illustration style was modified from (21)).

Procedure

Participants first read and signed an informed consent and eligibility was confirmed by a verbal interview. Participants were instructed on how to perform the navigation tasks, and were given an opportunity to practice entering an address for all three destination-entry modes while seated in the lab. Once participants were able to correctly enter an address using all three modes, they moved to the driving simulator where they were trained on how to perform the DRT, completed an introductory drive, and then practiced the dual task of driving and responding to the DRT.

The experimental period consisted of three counterbalanced blocks corresponding to each of the destination entry modes. Following the ISO guidelines (7), each experiment block began with a training period building-up to the triple task of driving, responding to the DRT, and entering a destination. Participants performed the destination entry task while stationary, first without, and then with, the DRT. The process was then repeated while driving. Participants were required to achieve proficiency on each training condition (defined as performing the task correctly while also responding to at least 70% of the DRT stimuli) before advancing to the next stage.

TABLE 1 Destination entry steps for all three modes.

Step	Voice: Standard	Voice: Hands-Free	Manual (Touchscreen)
1	Tap Home button to wake up screen		
2	Double-tap Home button		Open Google Maps application
3	Speak: “Navigate to 3-8-5 Prospect St, Cambridge”	Speak: “Hi Galaxy”	Tap Search bar
4		Speak: “Navigate to 3-8-5 Prospect St, Cambridge”	Type address: “385 Prospect St, Cambridge” and select address when it appears
5			Select the car icon to show routes
6			Select “Start navigation”

During the evaluation periods, participants first completed three minutes of single-task driving; the latter two minutes of this period were used to define the “Baseline Period.” Participants were then asked to enter the address “177 Massachusetts Ave, Cambridge” while simultaneously responding to the DRT. The task was considered complete once the participant stated that the device had finished calculating directions. Five seconds after completing the task, participants were asked to cancel the address by pressing the phone’s “Back” button until they reached the home screen. Participants then experienced a 30-second separation period, followed by a 60-second baseline DRT period, followed by another 30-second separation. They then completed a second destination entry (“293 Beacon St, Boston”) while responding to the DRT and then again returned to the home screen.

Glance Coding

Glances were coded manually from video, using a technique called dual coding with mediation (16). Two analysts independently manually coded glance locations from the in-vehicle video of the drivers’ faces using the MIT AgeLab Video Annotator (<https://bitbucket.org/agelab/annotator>). A third analyst mediated variation between the two analysts (e.g., regarding glance location, or timing onsets/offsets of more than 200 ms for a coded glance). Nine glance regions were coded: (a) road, (b) phone/device, (c) instrument cluster, (d) rearview mirror, (e) right, (f) left, (g) research assistant, (h) other, and (i) eyes not visible.

Data Reduction and Analysis

A linear mixed-effect model with task type as a fixed within-participant effect and participant as a random effect was applied with the Kenward-Roger correction to adjust the *F* statistics and degrees of freedom. As an effect size, both marginal R^2 , which describes the proportion of variance explained by the fixed factor alone, and conditional R^2 , which describes the proportion explained by the fixed and random factors, were reported. For post-hoc tests, paired t-test was applied to the DRT measurements (reaction time and miss rate) and Welch two sample t-test was applied to the buffer measurements (mean and standard deviation) due to unbalanced sample

sizes across conditions. Analyses were performed using R (31).

RESULTS

First, the ability of the DRT and AttenD buffer to capture a difference between visual-manual and auditory-vocal tasks was tested. Result showed that there was a significant effect of the task type on (a) DRT response time, $F(2, 42) = 5.16, p < .01$, marginal $R^2 = .07$, and conditional $R^2 = .57$, (b) buffer mean, $F(2, 39.46) = 211.12, p < .001$, marginal $R^2 = .85$, and conditional $R^2 = .88$, and (c) buffer SD, $F(2, 39.15) = 103.81, p < .001$, marginal $R^2 = .68$, and conditional $R^2 = .81$ (Figure 3 and Table 2). However, there was no significant effect of the task type on DRT percent missed. Post-hoc t-tests showed that there were significant differences between the visual-manual task and two auditory-vocal tasks for (a) DRT response time, $t(21) = 3.54, p < .01$ for the voice hands free and $t(21) = 2.23, p < .05$ for the voice standard, (b) buffer mean, $t(24.57) = -13.9, p < .001$ for the voice hands free and $t(26.22) = -13.47, p < .001$ for the voice standard, and (c) buffer SD, $t(36.19) = 10.71, p < .001$ for the voice hands free and $t(32.78) = 9.95, p < .001$ for the voice standard. However, there were no statistical differences between the two voice tasks for either the DRT or buffer. The results indicate that the manual tasks led to slower DRT response time, lower buffer mean, and higher buffer standard deviation (SD) compared to the two auditory-vocal tasks, and the magnitude of the effect was greater for the buffer measurements compared to DRT measurements.

TABLE 2 Summary of ANOVA tests.

	<i>F</i>	<i>df</i>	<i>p</i>	Marginal R^2	Conditional R^2
DRT RT	5.16	2, 42	0.01	0.07	0.57
Buffer mean	211.12	2, 39.46	0.001	0.85	0.88
Buffer SD	103.81	2, 39.15	0.001	0.68	0.81

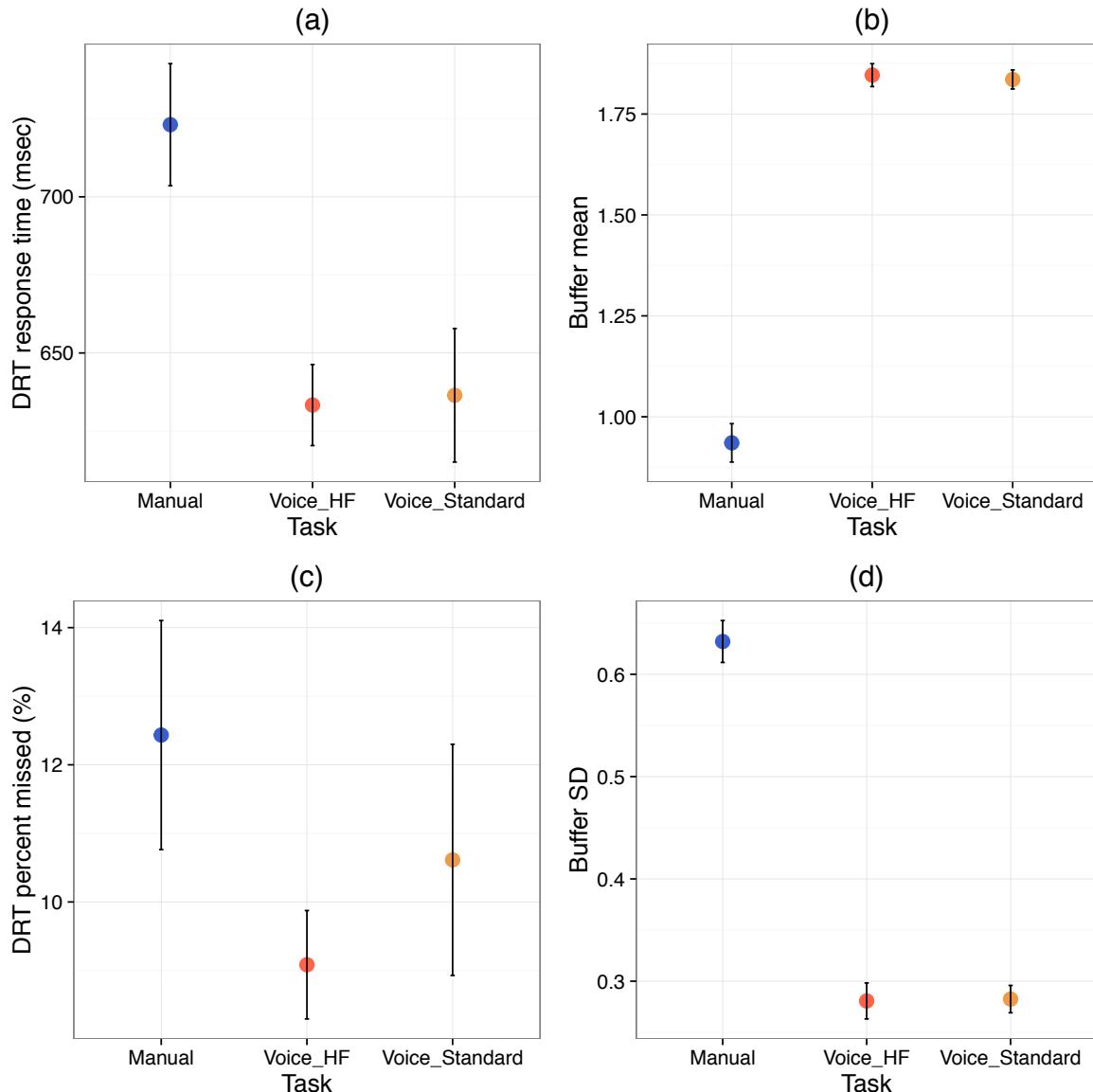


FIGURE 3 Task-level comparison of (a) DRT response time, (b) mean buffer, (c) DRT percent missed, and (d) SD buffer across task types (note: error bars indicate mean-adjusted standard error).

Second, a relationship between DRT measurements and buffer measurements was tested. Results showed that: (a) there was positive correlation between DRT response time and DRT percent missed, $r(60) = .49, p < .001$, indicating that slower responses were associated with higher miss rates and vice versa, (b) DRT response time was negatively correlated with buffer mean, $r(60) = -.3, p < .05$, and positively correlated with buffer SD, $r(60) = .25, p < .05$, and (c) DRT percentage of missed trials was positively correlated with buffer SD, $r(60) = .32, p < .05$ (Figure 4).

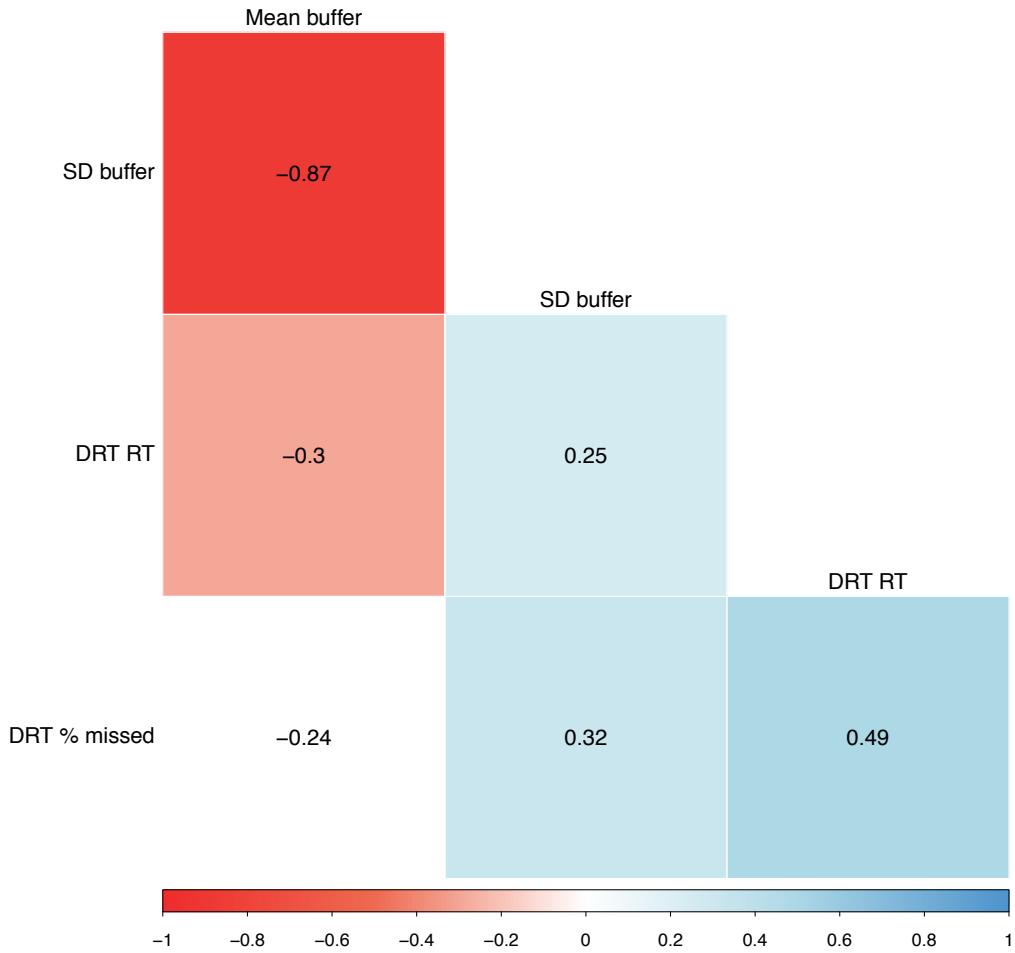


FIGURE 4 Correlation coefficients for all pairs between buffer and DRT measurements (note that colored cells represent statistically significant correlations).

DISCUSSION

A primary motivation for this paper was to examine how two different assessment methodologies - one presented as a measure of the effect of cognitive load on attention and the other, based on a consideration of how glances are allocated on and off the roadway and presented as a measure of attention – compare in how they rank the attentional demands of actual user interfaces in a driving context. This is of particular interest in that one of the HMIs evaluated represented what would traditionally be thought of as a primary visual-manual interface, while the other two represented slight variants on what might classically be considered as HMIs characterized by cognitive demand (because of their auditory-vocal aspects). As already discussed, the latter HMIs might more realistically be considered as multi-modal as some visual-manual demand characteristics were certainly present. Broadly speaking, the results show that both the remote DRT and the AttenD assessment methods produced results that would lead to similar conclusions concerning the relative attentional demand of the three HMIs, particularly if the DRT response time metric is given more weight than the miss percentage metric. More

specifically, the visual-manual tasks led to slower DRT response times and lower buffer mean values (more inattention), as compared to the auditory-vocal (multi-modal) tasks.

Looking more closely at the data, it appears that the AttenD algorithm was able to differentiate the visual-manual and auditory-vocal (multi-modal) interfaces (Figure 3) with greater confidence (i.e., the standard error was much tighter around the AttenD values and there was a larger effect size) than the DRT metrics. It can also be observed that the AttenD algorithm's mean and SD values were quite similar in their ability to differentiate the tasks.

The finding that the DRT miss percentage value did not prove as consistent in grouping and discriminating between the three HMIs as DRT response time is not an entirely unexpected result. Miss rates are generally relatively low for most real-world production HMIs used in the driving environment. Because of the limited number of DRT stimuli presented during the duration of such HMI tasks, a single miss can have a major impact on the effective miss percentage, resulting in a less stable measure than response time.

A positive correlation between DRT response time and DRT percent missed (slower responses times were associated with higher miss rates) was accompanied by a negative correlation between response time and mean buffer values, and a positive correlation between response time and buffer SD. The results may indicate that more attention to the roadway (i.e., higher buffer mean) leads to faster DRT responses, and greater variability of the forward attention (i.e., higher buffer SD) leads to slower DRT responses. The results show that DRT measurements are moderately correlated with the attention buffer measurements. While the magnitude of the correlations is only moderate, the fact that the relationships between a continuous measure (the AttenD algorithm) and a measure that probes attention only intermittently (DRT) are significant, is noteworthy.

Several implications arise from these data. First, they indicate that measures based on the allocation of visual attention may to some extent tap into the effects of cognitive demand associated with multi-tasking or at least provide effectively similar discrimination of relevant HMIs as provided by the remote DRT. There is a need for further research investigating connections between outputs from these two tools in the context of other multi-modal HMIs. At minimum, however, it is appropriate for investigations to consider a more unified theoretical approach to attention and its management during multitasking than the classical segregation of demand into visual, manual, and cognitive domains. Second, given that both methods differentiate between primary visual-manual and multi-modal HMIs, an AttenD style approach presents some strong advantages. The effect sizes seen in the present comparison effectively suggest that similar understanding of the differences between multi-modal HMIs could be obtained running fewer participants utilizing AttenD (or a conceptually related algorithm), providing a savings of time and money, or providing the freedom to run more comparisons when evaluating poorly understood, novel designs, such as Google Glass (1, 32). Moreover, constructing the types of attention buffer profiles presented here requires no artificial tertiary tasks, yielding potential benefits in terms of environmental validity. Future work may even be able to exploit this point to calculate the actual impact of tasks used in various forms of DRT investigation.

Further, an AttenD style approach to demand assessment may be retroactively generated from existing visual data. Many past efforts have already collected glance data of the type needed by an algorithm like AttenD (e.g., glance data are often acquired to assess visual demands under various voluntary guidelines (*for instance*, 33, 34)). Given that the AttenD profiles provide a continuous indication of drivers' attention/inattention, outputs can be utilized to conceptualize a more complete picture of attentional effects of HMI interaction. Finally, assessments can be easily expanded to on-road data where the collection of DRT metrics are

more difficult or infeasible. As with simulation data, AttenD style algorithms could be retroactively applied to naturalistic datasets. In this context, differences in the attentional characteristics of baseline, near-crash and crash events can be compared under the context of a framework balancing attentional demands to and from the road. Our efforts are currently focused on optimizing the sensitivity of more advanced implementations of the buffer concept for HMI evaluation and real-time driver state estimation. In this context, further refinement of the attention buffer concept is being developed to optimize the measurement to balance the safety relevance of visual and cognitive demands on driver attention.

ACKNOWLEDGEMENTS

Support for this work was provided by the Advanced Human Factors Evaluator for Automotive Demand (AHEAD) Consortium, the US DOT's Region I New England University Transportation Center at MIT, and the Toyota Class Action Settlement Safety Research and Education Program. The views and conclusions being expressed are those of the authors, and have not been sponsored, approved, or endorsed by Toyota or plaintiffs' class counsel.

REFERENCES

1. Sawyer, B. D., V. S. Finomore, and A. A. Calvo. Google Glass: A driver distraction cause or cure? *Human Factors: The Journal of the Human Factors and Ergonomic Society*, Vol. 57, No. 7, 2014, pp. 1307–1321.
2. He, J., J. S. McCarley, and A. F. Kramer. Lane keeping under cognitive load performance changes and mechanisms. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 56, No. 2, 2014, pp. 414–426.
3. Törnros, J., and A. Bolling. Mobile phone use – effects of conversation on mental workload and driving speed in rural and urban environments. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 9, No. 4, 2006, pp. 298–306.
4. Hancock, P. A., and J. S. Warm. A dynamic model of stress and sustained attention. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 31, No. 5, 1989, pp. 519–537.
5. Sawyer, B. D., A. A. Calvo, and V. S. Finomore. Serendipity in simulation: Building environmentally valid driving distraction evaluations of Google Glass™ and an Android™ smartphone. *Presented at the 19th Triennial Congress of the International Ergonomics Association*, 2015.
6. Harbluk, J. L., and P. C. Burns. Detection response tasks: Using remote, headmounted and Tactile signals to assess cognitive demand while driving. *Presented at the 7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, 2013.
7. International Organization for Standardization. *ISO 17488 Road vehicles—Transport information and control systems—Detection-Response Task (DRT) for assessing attentional effects of cognitive load in driving*. ISO/DIS, Geneva, Switzerland, 2013.
8. Ranney, T. A., G. H. S. Baldwin, L. A. Smith, E. N. Mazzae, and R. S. Pierce. *Detection Response Task (DRT) Evaluation for Driver Distraction Measurement Application (DOT HS 812 077)*. Publication DOT HS 812 077. National Highway Traffic Safety Administration, Washington, DC, 2014.
9. Merat, N., G. Kountouriotis, O. Carsten, and J. Engström. Performance on the detection response task during driving: Separating the manual and cognitive element of the secondary task. *Presented at the 4th International Driver Distraction and Inattention*

- Conference*, 2015.
10. Angell, L., J. Auflick, P. A. Autria, D. Kochhar, L. Tijerina, W. Biever, T. Diptiman, J. Hogsett, and S. Kiger. *Driver Workload Metrics Project Task 2 Final Report*. Publication DOT HS 810 635. National Highway Traffic Safety Administration, Washington, D.C., 2006.
 11. Carter, C., and R. Graham. Experimental comparison of manual and voice controls for the operation of in-vehicle systems. *Presented at the Human Factors and Ergonomic Society Annual Meeting*, 2000.
 12. Dobres, J., B. Reimer, B. Mehler, J. Foley, K. Ebe, B. D. Seppelt, and L. Angell. The influence of driver's age on glance allocation during single-task driving and voice vs. visual-manual radio tuning (No. 2016-01-1445). *SAE Technical Paper*, 2016.
 13. Chiang, D. P., A. M. Brooks, and D. H. Weir. On the highway measures of driver glance behavior with an example automobile navigation system. *Applied Ergonomics*, Vol. 35, No. 3, 2004, pp. 215–223.
 14. Mehler, B., D. Kidd, B. Reimer, I. Reagan, J. Dobres, and A. McCartt. Multi-modal assessment of on-road demand of voice and manual phone calling and voice navigation entry across two embedded vehicle systems. *Ergonomics*, Vol. 59, No. 3, 2016, pp. 344–367.
 15. Lee, J. D., B. Caven, S. Haake, and T. L. Brown. Speech-Based Interaction with In-Vehicle Computers: The Effect of Speech-Based E-Mail on Drivers' Attention to the Roadway. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 43, No. 4, 2001, pp. 631–640.
 16. Reimer, B., B. Mehler, J. Dobres, and J. F. Coughlin. *The Effects of a Production Level "Voice-Command" Interface on Driver Behavior: Summary Findings on Reported Workload, Physiology, Visual Attention, and Driving Performance*. Publication 2013-17A. MIT AgeLab, Cambridge, MA, 2013.
 17. Strayer, D. L., J. M. Cooper, J. Turrill, J. Coleman, N. Medeiros-Ward, and F. Biondi. *Measuring Cognitive Distraction in the Automobile*. AAA Foundation for Traffic Safety, Washington, DC, 2013.
 18. Strayer, D. L., J. Turrill, J. Coleman, E. V. Ortiz, and J. M. Cooper. *Measuring Cognitive Distraction in the Automobile II: Assessing In-Vehicle Voice-Based Interactive Technologies*. AAA Foundation for Traffic Safety, Washington, DC, 2014.
 19. Strayer, D. L., J. M. Cooper, J. Turrill, J. Coleman, and R. J. Hopman. *Measuring Cognitive Distraction in the Automobile III: A Comparison of Ten 2015 In-Vehicle Information Systems*. AAA Foundation for Traffic Safety, Washington, DC, 2015.
 20. Bruyas, M. P., L. Dumont, and F. Bron. Sensitivity of Detection Response Task (DRT) to the driving demand and task difficulty. *Presented at the 7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, 2013.
 21. Kircher, K., and C. Ahlström. Issues related to the driver distraction detection algorithm AttenD. *Presented at the 7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Gothenburg, Sweden, 2009.
 22. Kircher, K., C. Ahlström, and A. Kircher. Comparison of two eye-gaze based real-time driver distraction detection algorithms in a small-scale field operational test. *Presented at the 7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, 2009.
 23. Kircher, K., and C. Ahlström. *The Driver Distraction Detection Algorithm AttenD*. In Driver Distraction and Inattention (M. Regan, J. D. Lee, and T. W. Victor, eds.), Farnham, UK, pp. 327–348, 2013.

24. Lee, J., J. Moeckli, T. Brown, S. Roberts, T. Victor, D. Marshall, C. Schwarz, and E. Nadler. Detection of driver distraction using vision-based algorithms. *Presented at the 23rd International Technical Conference on the Enhanced Safety of Vehicles*, 2013.
25. Reimer, B., B. Mehler, J. Dobres, H. McAnulty, A. Mehler, D. Munger, and A. Rumpold. Effects of an “expert mode” voice command system on task performance, glance behavior & driver physiology. *Presented at the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, Seattle, WA, 2014.
26. Mehler, B., B. Reimer, J. Dobres, J. Foley, and K. Ebe. Additional findings on the multi-modal demands of “voice-command” interfaces (2016-01-1428). *SAE Technical Paper*, 2016.
27. Reimer, B., B. Mehler, I. Reagan, D. Kidd, and J. Dobres. Multi-modal demands of a smartphone used to place calls and enter addresses during highway driving relative to two embedded systems. *Ergonomics*, 2016, pp. 1–21.
28. Munger, D., B. Mehler, B. Reimer, J. Dobres, A. Pettinato, B. Pugh, and J. F. Coughlin. A simulation study examining smartphone destination entry while driving. *Presented at the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, Seattle, WA, 2014.
29. Reimer, B., and B. Mehler. The impact of cognitive workload on physiological arousal in young adult drivers: a field study and simulation validation. *Ergonomics*, Vol. 54, No. 10, 2011, pp. 932–942.
30. Wang, Y., B. Mehler, B. Reimer, V. Lammers, L. A. D'Ambrosio, and J. F. Coughlin. The validity of driving simulation for assessing differences between in-vehicle informational interfaces: A comparison with field testing. *Ergonomics*, Vol. 53, No. 3, 2010, pp. 404–420.
31. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. 2008.
32. Beckers, N., S. Schreiner, P. Bertrand, B. Mehler, and B. Reimer. Comparing the demands of destination entry using Google Glass and the Samsung Galaxy S4 during simulated driving. *Applied Ergonomics*, Vol. 58, 2017, pp. 25–34.
33. National Highway Traffic Safety Administration. *Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices*. Publication NHTSA-2010-0053. National Highway Traffic Safety Administration, Washinton, D.C., 2013.
34. Driver Focus-Telematics Working Group. *Statement of Principles, Criteria, and Verification Procedures on Driver-Interactions with Advanced In-Vehicle Information and Communication Systems*. Alliance of Automobile Manufactures, Washington, D.C., 2006.