

Linking the Detection Response Task and the AttenD Algorithm Through Assessment of Human–Machine Interface Workload

Joonbum Lee, Ben D. Sawyer, Bruce Mehler, Linda Angell, Bobbie D. Seppelt, Sean Seaman, Lex Fridman, and Bryan Reimer

Multitasking related demands can adversely affect drivers' allocation of attention to the roadway, resulting in delays or missed responses to roadway threats and to decrements in driving performance. Robust methods for obtaining evidence and data about demands on and decrements in the allocation of driver attention are needed as input for design, training, and policy. The detection response task (DRT) is a commonly used method (ISO 17488) 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. A relationship between the two tools is explored. A previous multitasking driving simulation study, which used the remote form of the DRT to differentiate the demands of a primary visual–manual human–machine interface from alternative primary auditory–vocal multimodal human–machine interfaces, 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 human–machine interfaces 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 analyses of past and future data sets, and possible application for driver state detection are also discussed.

Operating a vehicle is a task with great variability in demand owing 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 multitasking, and either driving or in-vehicle task performance, or both, may degrade (1). Drivers commonly attempt to compensate for increased demand from multitasking, 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 owing 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, several approaches have been taken. Peripheral detection tasks (PDTs) were succeeded by detection response tasks (DRTs) (6), which have been characterized in recent years as a method for measuring the attentional effects of cognitive load. Methods for use of DRT as a cognitive load measure are described within ISO 17488 [Road Vehicles—Transport Information and Control Systems—Detection-Response Task (DRT) for Assessing Attentional Effects of Cognitive Load in Driving, 2013]. 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 s (measured from one stimulus onset to the next), and are sustained for an exposure duration of 1 s. The DRT task is designed to be presented concurrently while the driver performs other, secondary tasks. Measurements of response time and accuracy (hit and 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, 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 under way. 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

J. Lee and L. Fridman, E40-215; B.D. Sawyer, B. Mehler, and B. Reimer, E40-279, MIT AgeLab and New England University Transportation Center, 77 Massachusetts Avenue, Cambridge, MA 02139. L. Angell, B.D. Seppelt, and S. Seaman, Touchstone Evaluations, Inc., 440 Burroughs Street, Detroit, MI 48202. Corresponding author: J. Lee, joonbum@mit.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2663, 2017, pp. 82–89.
<http://dx.doi.org/10.3141/2663-11>

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., 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 [e.g., Ranney et al. (7)]. Signal detection analysis approaches can help to refine such analyses by disambiguating the accuracy measure (8).

The value of such data is multifaceted, but one promising use is in the evaluation of the demands associated with in-vehicle user interfaces (9). 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 (10–13). However, the extent to which elevated cognitive demand is present in such interactions has been raised (14–16). One body of work (16–18) 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 [e.g., Ranney et al. (7)]. 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. Furthermore, the introduction of the response task itself is the introduction of further multitasking, 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 observed that mental workload was rated higher in simulated driving with visual or tactile DRTs than without (19). 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 uninterrupted 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 and Kircher et al. introduced the AttenD algorithm to detect distracted driving based on the allocation of visual resources (20–22). 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 nondriving-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 compared with other vision-based distraction detection algorithms (23). The AttenD algorithm provides a continuous measure over an epoch of driving or multitasking, which is important.

Kircher and Ahlström and Kircher et al. explicitly presented the AttenD algorithm as a method to provide a continuous indication of the “extent of inattention or distraction” shown by a driver (20–22). While it is not the primary focus of the current paper, those 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 narrower 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 powerful tool in understanding the impact of in-vehicle HMI on driver attention.

A key research question the present work addresses is this: 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). Control and selection modalities now extend beyond manual manipulation to frequently include voice-command options. The traditional division of demand into silos of visual, manual, and cognitive resources may not be the most useful approach given the reality of modern in-vehicle HMIs. Cognitive resources are arguably involved in all HMI interactions, and those production interfaces characterized as cognitive (e.g., auditory–vocal) might better be considered as multimodal in that they often involve significant draw on visual resources (24–26). 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 reanalyze glance data from a simulator study on driver multitasking that used the remote DRT technique. Broadly, the researchers 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 across several HMIs. In comparing the two assessment approaches, this data source supported consideration of a classically defined visual–manual HMI interaction as compared with two variants of an auditory–vocal HMI for completing the same secondary task goal.

METHODS

This study is a secondary analysis of data from Munger et al. (27), 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 are summarized as follows.



(a)



(b)

FIGURE 1 (a) MIT AgeLab driving simulator at time of study and (b) location of the remote visual DRT light on the lower windshield in line of sight of the forward roadway.

Participants

Participants were recruited across two age groups (20 to 24 and over 55) from the greater Boston area using online and newspaper postings. Participants were required to meet several criteria: (a) a valid driver's license for more than 3 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. Of the 24 participants from the original study analysis set, 2 cases had video image issues that precluded coding of the glance data, which resulted in 22 cases being available for the present analysis (mean age = 46.18, SD = 21.51, min = 20, max = 68, $n = 9$ females and 13 males).

Apparatus

The study used 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, California) based on a driver's interaction with the steering wheel, brake, and accelerator. Instructions and audio tasks were prerecorded and presented through the vehicle sound system. Correspondence between the demands of this simulator con-

figuration and actual driving scenarios has been established through previous research (28, 29). 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 ISO 17488. A red LED was mounted on the windshield within the field of view of the forward roadway, and responses were recorded from a microswitch placed on the participant's left index finger. Following the standard, the LED was activated every 3 to 5 s for a period of 1 s or until the participant responded using the finger-mounted switch.

Destination Entry Device

Participants entered destination addresses in a Samsung Galaxy S4 (model number SCH-1545; released March 2013) with a 5-in. display and $1,920 \times 1,080$ resolution. The device was free floating (not mounted), and participants held it in their hands or rested it on the center console or other location at their discretion while performing tasks. Navigation tasks were carried out using Google Maps. Participants were extensively trained on how to enter an address using three interaction modes (Table 1). One mode required visual-manual touchscreen interaction; the two other modes were auditory-vocal-visual-manual (e.g., voice-based commands). For all three methods, participants had to first wake up the phone by pressing the large home button at the bottom of the screen. To enter voice recognition mode, participants double-tapped the home button. For the touch interface, participants opened Google Maps, typed a specified address

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". |

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 home button. The phone then presented one of several possible introductions, for example, “Hello, I hope you are having a great and productive week. If you need any help say, ‘Hi Galaxy’.” After saying “Hi Galaxy,” the command structure was “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-s time buffer, which increases while a driver looks on-road and decreases while the driver looks away from the road. The closer the buffer falls to zero, the closer the driver is assumed to be to a distracted state. More details of the algorithm can be found in Kircher and Ahlström’s and Kircher et al.’s original manuscripts (20–22), which were used to perform buffer calculations. Figure 2 illustrates one participant’s buffer profile during the manual destination entry task.

Procedure

Participants read and signed an informed consent on arrival. Participants were then instructed on how to perform the navigation tasks, and they were given an opportunity to practice entering an address for all three destination-entry modes on a tablet 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 ISO guidelines, each experimental 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.

During evaluation periods, participants engaged in 3 min of single-task driving and were then asked to enter the address “177 Massachusetts Ave, Cambridge” while simultaneously responding to the DRT. After the device had finished calculating directions, participants canceled the address by pressing the phone’s “Back” button until they reached the home screen. After a 30-s separation period, a 60-s baseline DRT period followed, and then another 30-s separation. Participants then completed a second destination entry (“293 Beacon St, Boston”) while responding to the DRT.

Glance Coding

Glances were coded manually from video of drivers’ faces. Following recommendations in Reimer et al. (15), two analysts independently coded glance locations using the MIT AgeLab video annotator (<https://bitbucket.org/agelab/annotator>), and a third analyst mediated variation (e.g., in regard to glance location, or timing onsets and offsets of more than 200 ms for a coded glance). Nine glance regions were coded: (a) road, (b) phone and 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, a paired t -test was applied to the DRT measurements (reaction time and miss rate), and a Welch two sample t -test was applied to the buffer measurements (mean and standard deviation) owing to unbalanced sample sizes across conditions. Analyses were performed using R (30).

RESULTS

First, the ability of the DRT and AttenD buffer to capture a difference between visual–manual and auditory–vocal tasks was tested. Results showed that there was a significant effect of 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, $F(2, 39.46) = 211.12, p < .001$,

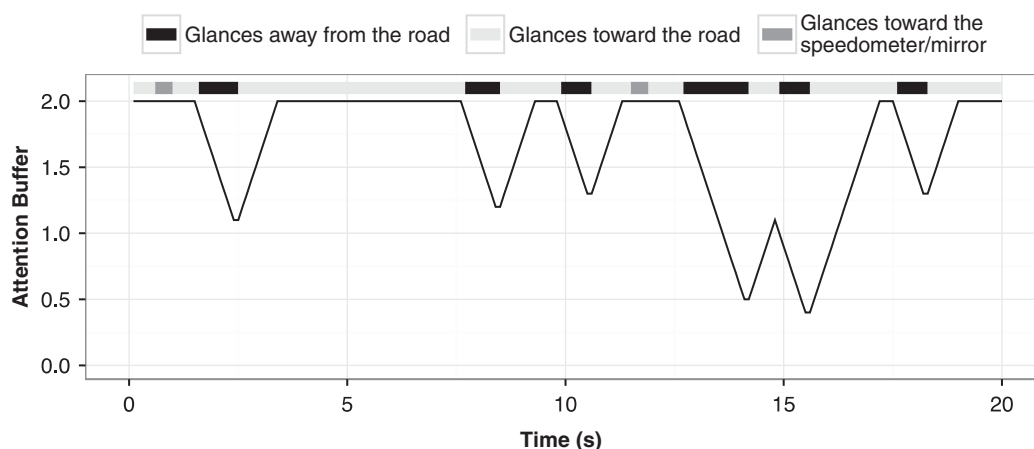


FIGURE 2 Example of a buffer profile over time; illustration style was modified from Kircher and Ahlström (20).

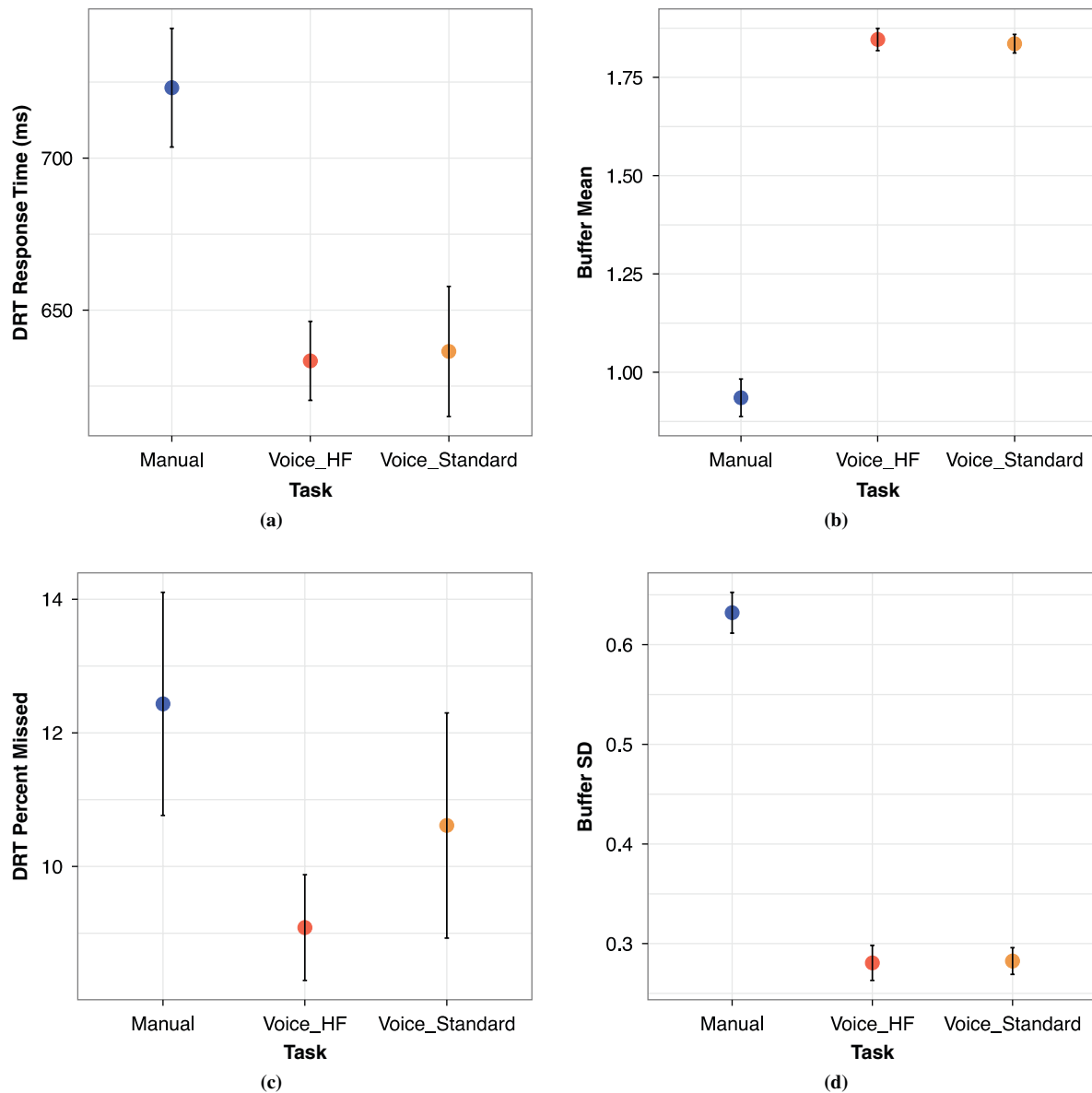


FIGURE 3 Task-level comparison of (a) DRT response time, (b) mean buffer, (c) DRT percent missed, and (d) SD buffer across task types; error bars indicate mean-adjusted standard error.

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 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 with the two auditory-vocal tasks (Figure 4); the magnitude of the effect was greater for the buffer measurements compared to DRT measurements.

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

TABLE 2 Summary of Analysis of Variance Tests

| | <i>F</i> | <i>df</i> | <i>p</i> | Marginal R^2 | Conditional R^2 |
|-------------|----------|-----------|----------|----------------|-------------------|
| DRT RT | 5.16 | 2, 42 | .01 | .07 | .57 |
| Buffer mean | 211.12 | 2, 39.46 | .001 | .85 | .88 |
| Buffer SD | 103.81 | 2, 39.15 | .001 | .68 | .81 |

NOTE: RT = response time; *df* = degrees of freedom.

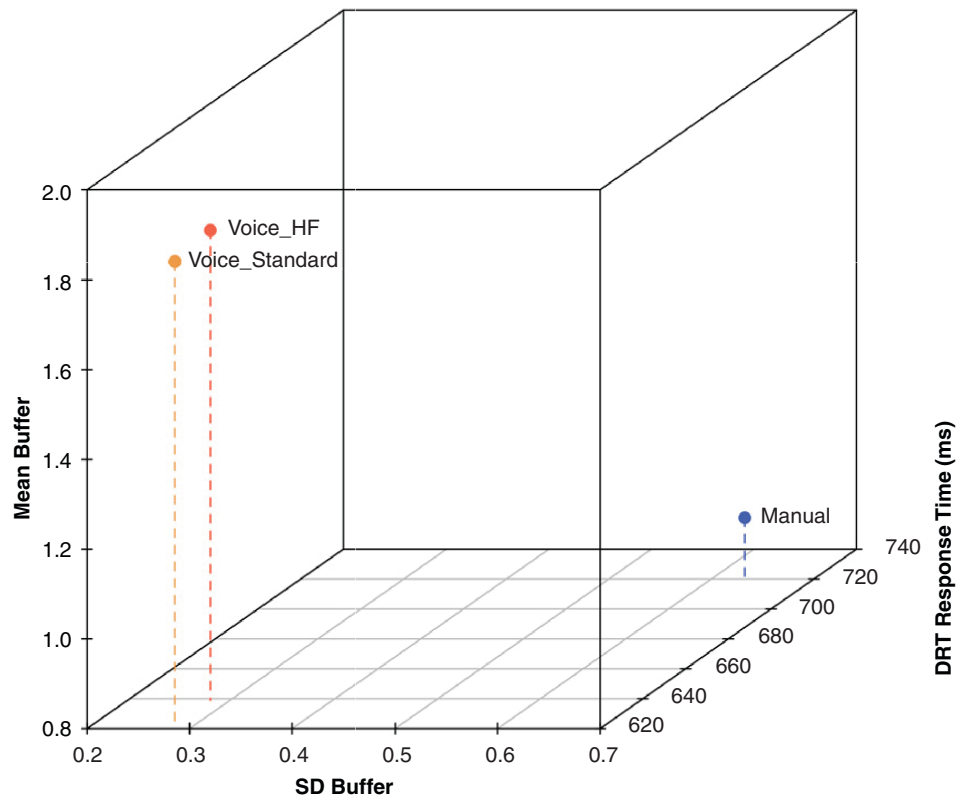


FIGURE 4 Task-level three-dimensional differentiation of voice-based and visual-manual HMIs in regard to mean and standard deviation of attention buffer values and DRT response time.

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 5).

DISCUSSION OF RESULTS

A primary motivation for this study 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, 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 multimodal since 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 with the auditory-vocal (multimodal) tasks.

Looking more closely at the data, it appears that the AttenD algorithm was able to differentiate the visual-manual and auditory-vocal (multimodal) 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 negative correlations between DRT measures and mean buffer values, and positive correlations between DRT measures 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 forward attention (i.e., higher buffer SD) leads to slower DRT responses. 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 (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 multitasking or at least provide effectively similar discrimination of relevant HMIs as provided by the remote DRT. There is a need

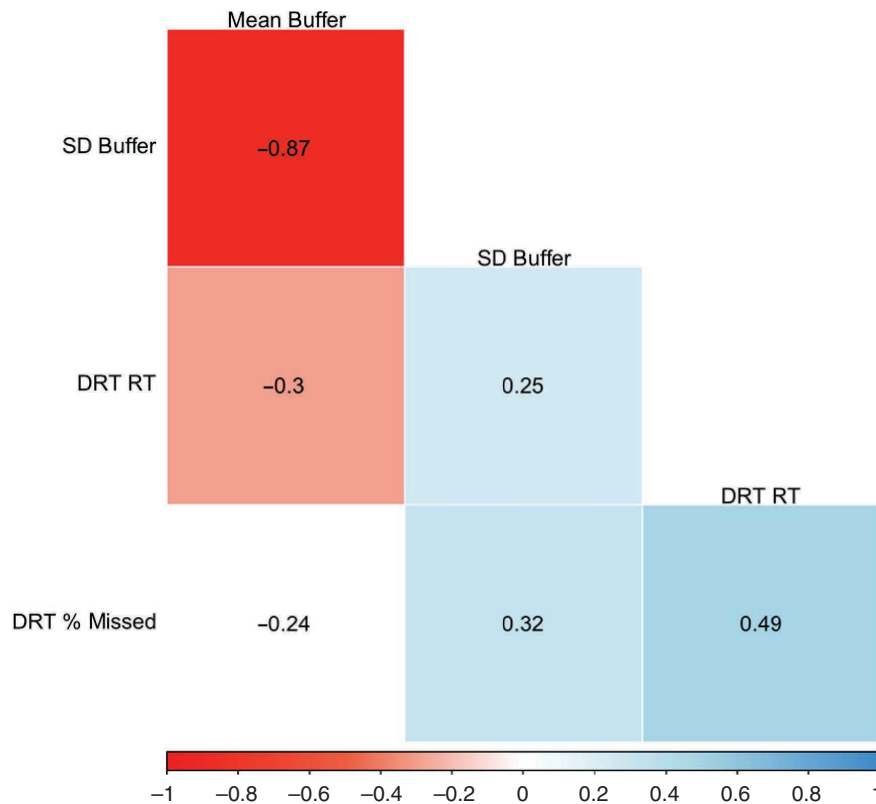


FIGURE 5 Correlation coefficients for all pairs between buffer and DRT measurements; colored cells represent statistically significant correlations.

for further research investigating connections between outputs from these two tools in the context of other multimodal 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 multimodal 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 multimodal HMIs could be obtained running fewer participants using 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, 31). 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.

Furthermore, 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; for example, glance data are often acquired to assess visual demands under various voluntary guidelines [e.g., NHTSA (32) and Alliance of Automobile Manufacturers (33)]. Given that the AttenD profiles provide a continuous indication of drivers' attention and inattention, outputs can be used 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 is more difficult or infeasible. As with simulation data, AttenD style algorithms could be retroactively applied

to naturalistic data sets [e.g., Seaman et al. (34)]. 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. The researchers' 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.

ACKNOWLEDGMENTS

Support for this work was provided by the Advanced Human Factors Evaluator for Automotive Demand Consortium, the U.S. Department of Transportation's Region 1 New England University Transportation Center at MIT, and the Toyota Class Action Settlement Safety Research and Education Program.

REFERENCES

1. Sawyer, B.D., V.S. Finomore, A.A. Calvo, and P.A. Hancock. Google Glass: A Driver Distraction Cause or Cure? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 56, No. 7, 2014, pp. 1307–1321. <https://doi.org/10.1177/0018720814555723>.
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. <https://doi.org/10.1177/0018720813485978>.
3. Törmös, J., and A. Bolling. Mobile Phone Use—Effects of Conversation on Mental Workload and Driving Speed in Rural and Urban Envi-

- ronments. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 9, No. 4, 2006, pp. 298–306. <https://doi.org/10.1016/j.trf.2006.01.008>.
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 19th Triennial Congress of the International Ergonomics Association, Melbourne, Victoria, Australia, Aug. 9–14, 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 7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, Bolton Landing, N.Y., June 17–20, 2013.
 7. 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*. Publication DOT HS 812 077. NHTSA, U.S. Department of Transportation, 2014.
 8. 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 4th International Driver Distraction and Inattention Conference, Sydney, New South Wales, Australia, Nov. 9–11, 2015.
 9. 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. NHTSA, U.S. Department of Transportation, 2006.
 10. Carter, C., and R. Graham. Experimental Comparison of Manual and Voice Controls for the Operation of In-Vehicle Systems. Presented at Human Factors and Ergonomics Society Annual Meeting, San Diego, Calif., July 29–Aug. 4, 2000. <https://doi.org/10.1177/154193120004402016>.
 11. 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. SAE Technical Paper 2016-01-1445. SAE, 2016.
 12. 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. <https://doi.org/10.1016/j.apergo.2004.01.005>.
 13. 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. <https://doi.org/10.1080/00140139.2015.1081412>.
 14. 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. <https://doi.org/10.1518/001872001775870340>.
 15. 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, Mass., 2013.
 16. 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, D.C., 2013.
 17. 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, D.C., 2014.
 18. 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, D.C., 2015.
 19. Bruyas, M.P., L. Dumont, and F. Bron. Sensitivity of Detection Response Task (DRT) to the Driving Demand and Task Difficulty. Presented at 7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, Bolton Landing, N.Y., June 17–20, 2013.
 20. Kircher, K., and C. Ahlström. Issues Related to the Driver Distraction Detection Algorithm AttenD. Presented at 1st International Conference on Driver Distraction and Inattention, Gothenburg, Sweden, Sept. 28–29, 2009.
 21. 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. In *Proceedings of the 5th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Big Sky, Mont., June 22–25, 2009.
 22. 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.), Ashgate, Farnham, United Kingdom, 2013, pp. 327–348.
 23. 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 23rd International Technical Conference on the Enhanced Safety of Vehicles, Seoul, South Korea, June 25–28, 2013.
 24. 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 and Driver Physiology. Presented at 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Seattle, Wash., Sept. 17–19, 2014. <https://doi.org/10.1145/2667317.2667320>.
 25. Mehler, B., B. Reimer, J. Dobres, J. Foley, and K. Ebe. Additional Findings on the Multi-Modal Demands of "Voice-Command" Interfaces. SAE Technical Paper No. 2016-01-1428. SAE, 2016.
 26. 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*, Vol. 59, No. 12, 2016, pp. 1565–1585. <https://doi.org/10.1080/00140139.2016.1154189>.
 27. 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 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Seattle, Wash., Sept. 17–19, 2014. <https://doi.org/10.1145/2667317.2667349>.
 28. 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. <https://doi.org/10.1080/00140139.2011.604431>.
 29. 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. <https://doi.org/10.1080/00140130903464358>.
 30. R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008.
 31. 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. <https://doi.org/10.1016/j.apergo.2016.05.005>.
 32. *Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices*. Publication NHTSA-2010-0053. NHTSA, U.S. Department of Transportation, 2013.
 33. 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 Manufacturers, Washington, D.C., 2006.
 34. Seaman, S., J. Lee, B. Seppelt, L. Angell, B. Mehler, and B. Reimer. It's All in the Timing: Using the AttenD Algorithm to Assess Texting in the NEST Naturalistic Driving Database. In *Proceedings of the 9th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Manchester Village, Vt., June 26–29, 2017.

The views and conclusions expressed in this paper are those of the authors and have not been sponsored, approved, or endorsed by Toyota or plaintiffs' class counsel.

The Standing Committee on Simulation and Measurement of Vehicle and Operator Performance peer-reviewed this paper.