Differentiating Cognitive Load Using a Modified Version of AttenD

Bobbie Seppelt
MIT AgeLab &
Touchstone
Evaluations, Inc.
Cambridge, US
bseppelt@mit.edu

Sean Seaman
Touchstone
Evaluations, Inc.
Grosse Pointe, US
sseaman@touchst
oneevaluations.co
m

Linda Angell
Touchstone
Evaluations, Inc.
Grosse Pointe, US
langell@touchsto
neevaluations.com

Bruce Mehler
MIT AgeLab
Cambridge, US
bmehler@mit.edu

Bryan Reimer
MIT AgeLab
Cambridge, US
reimer@mit.edu

ABSTRACT

Voice interfaces offer promise in allowing drivers to keep their eyes on-road and hands on-wheel. In relieving visual-manual demand, there is the potential for voice-enabled interfaces to inadvertently shift the burden of load to cognitive resources. Measurement approaches are needed that can identify when and to what extent cognitive load is present during driving. A modified form of the AttenD algorithm was applied to assess the amount of cognitive load present in a set of auditory-vocal task interactions. These tasks were subset from a larger on-road study conducted in the Boston area of driver response during use of an in-vehicle voice system [22]. The modified algorithm differentiated among the set of auditory-vocal tasks examined – and may be useful to HMI practitioners who are working to develop and evaluate HMIs to support drivers in managing their attention to the road, and in the development of real-time driver attention monitoring systems.

Author Keywords

Cognitive Load; Voice Interface; AttenD.

CCS Concepts

H.1.2 User/Machine Systems (e.g. Human factors); H.5.2 Voice I/O.

INTRODUCTION

In-vehicle "infotainment" and navigation features increasingly include voice-enabled functionality that interfaces with dynamic visual displays. These displays are commonly embedded in the center stack or instrument cluster, forming highly complex mixed-mode interactions that draw upon auditory, vocal, visual, manual and cognitive demands [24]. Voice interfaces offer promise in allowing

drivers to keep their eyes on the road and hands on the steering wheel. Intended by design to reduce competing visual-manual demands that can pull eyes and hands off-road, production-level voice command interfaces can also impose demands that undermine intended affordances for increased attention to driving [22]. How voice components are designed, developed, and integrated in-vehicle all influence the degree to which drivers' attentional resources are demanded [21,23] – from pacing of task structure, to number and complexity of input task commands, to presence and type of visual information, to the quality and responsiveness of embedded voice recognition. Beyond learned design experience, formal processes for evaluating how these attributes are balanced to promote effective attention management profiles (situation awareness to the road, etc.) has yet to be establish. Use of design approaches aimed at meeting existing visual-manual guidelines [e.g., 17] may, if temporal task structure is not assessed, negatively bias the effective and safe development of voice-enabled applications.

Dependent on implementation, there is the potential for voice-enabled interfaces to inadvertently shift the burden of load from visual-manual resources to cognitive resources (e.g., memory requirements for complex sets of voice commands, or retention of multiple pieces of information during address entry). For over a decade, the human factors community has debated the effects of cognitive load on driving performance and its relationship to road safety [e.g., 15,28,32] – and struggled to measure it. Central to this debate is the concern that certain nonvisual secondary activities may impose undue cognitive load that induces the withdrawal of attention from the driving task or "mind off road" [33], and how changes in visual scanning impact object and event detection [e.g., 14,16]. Considerable research on in-vehicle use of phones have drawn disparate conclusions on the extent to which cognitive distraction from conversation contributes to drivers failing to notice potential hazards that places them at increased risk of crash [e.g., 4,6,10,13].

A recent review [5] of the effects of cognitive load on driving performance summarized the broad experimental literature in this area with the tenet that certain types and magnitudes of cognitive load may selectively impair elements of driving.

In particular, secondary activities that invoke significant demands on executive control – those imposing working memory load, long-term memory retrievals, or task scheduling or information rehearsal – were identified to impair driving tasks that duly rely on cognitive control, such as object and event detection. Unaffected driving tasks include those that are automatized from routine, consistent performance. Highly-practiced elements of driving such as lane keeping or lateral vehicle control fit into this distinction due to robust sensory-motor mappings that have developed over a range of driving environments. This assertion for when cognitive load impairs driving performance is consistent with overarching findings that have differentially shown an afforded protective effect of cognitive load on driving [27,29] and a reduced naturalistic crash risk [8,18,34], to those that have shown its potential to induce detrimental effects to driving performance [31,32].

In addition to amount, various types of cognitive load have been studied to assess their safety-degrading effects on driving performance. As a form of secondary task-induced cognitive load, cell phone conversations, in particular, have received significant attention due to legislative concerns over the trend of increased phone use while driving. Researchers have studied listening and speaking aspects of conversation for both hands-free and hand-held implementations and have reached divergent conclusions on the risk of various implementations [2,20,31,34]. In part, these differences depend on selected baselines for calculating risk metrics and on the realism of selected tasks. Recent findings have shown that even in the absence of secondary tasks, drivers sometimes allocate their attention to cognitive activities such as mind wandering in various amounts during routine driving [e.g., 1,7] – perhaps in an effort to increase their own arousal and performance levels under conditions of monotony or boredom in the driving task or driving scene. In assessing how different types of secondary task load impact driver attention, it is important to consider measured effects relative to reference tasks that contain naturalistic amounts and types of cognitive load.

Stepping outside of the debate over the degree to which cognitive load does or does not pose a risk to driving safety, measurement approaches are needed that can identify when and to what extent it is present during driving. It has long been understood that increases in cognitive load result in a more concentrated focus on road center [6,28,35]. Measurement approaches such as the detection response task (DRT) have been interpreted as showing the effects of cognitive load [e.g., 19]. This measure, however, has specific input modality characteristics that can interact with the secondary task effects under study, creating a triple task assessment scenario. A different approach, based on the AttenD algorithm, has recently demonstrated its potential to differentiate between a range of mixed-mode interfaces characteristics with more power than analyses of DRT response times and miss rate, and without the potential for introducing implementation-specific resource demands [12].

The present work builds on the highlighted potential of the AttenD algorithm to differentiate visual-manual and auditory-vocal HMI demand types, and presents findings applying a modified form of the AttenD algorithm to a set of auditory-vocal task interactions that introduce cognitive load. These tasks were subset from a larger study of driver response in use of an in-vehicle voice system [22]. Specifically, three levels of cognitive load were examined within the context of a common in-vehicle task: listening to, thinking about, and retelling a story. In this work, data were extracted from a listening task, a speech production task (retelling a story), and a mental reflection task (thinking about the story). Prior research has shown that speech production imposes higher cognitive load than listening tasks [e.g., 3,9,30]. There is relatively little prior research on mental reflection tasks [25], but from this work, this purely cognitive task is expected to impose lower load relative to speech production and listening tasks. It was hypothesized that the AttenD algorithm used here would be sensitive to the level of cognitive load present in the set of extracted tasks.

METHODS

This study is a secondary analysis of data collected on-road to examine driver interaction with an on-board voice-command (SYNC™) system [22]. In this study, glance behavior was analyzed during periods of single-task driving and during concurrent performance for a set of visual-manual and auditory-vocal equivalent tasks performed using the factory-installed interface. The full methodological details can be found in the text and appendix of the initial paper. Key task details related to the subset epochs are summarized next.

Participants

Participants were recruited across two age groups (20-29 and 60-69) from the greater Boston area. Participants were required to meet a number of inclusion and exclusion criteria (see [22] for details) to ensure they were safe active drivers who were native English speakers. Upon arrival, participants were provided informed consent consistent with guidelines set forth by the Massachusetts Institute of Technology's Institutional Review Board as well as an overview of the study. Compensation was $90. Out of the total 75 participants from the original study analysis set, 19 cases were removed due to missing video or questionnaire data or from reported instances of malfunctioning or incomplete system interactions, resulting in 56 cases being available for the primary analysis in this study (mean age = 43.5, SD = 20.7, min = 20, max = 69, n of female participants = 27, and n of male participants = 29).

Apparatus

A 2010 Lincoln MKS with a factory installed voice-command system (SYNC™ for voice control of the USB-connected phone and media device and the Sirius Travel Link navigation system) was used. The vehicle encompasses a fairly typical interface with a large center-mounted display,

a radio volume knob situated just below the display, and a "push-to-talk" button located on the steering wheel. When active, the voice control interface provided supplemental visual information in support of driver-system dialog on the display screen located in the center stack. A cellular phone was paired to the vehicle, allowing drivers to make phone calls to a remote answering machine.

During on-road data collection, an experimenter was seated in the rear of the vehicle and was responsible for providing participants with driving directions, verifying participants followed instructions to complete tasks according to a predefined script, and ensuring safe vehicle operation. An on-board data acquisition system was used to play recorded audio for primary instructions and task descriptions via keyboard-enabled commands that the experimenter triggered at predefined points, thereby enabling consistent presentation of the task set along a specified route across participants. The data acquisition system recorded vehicle telemetry and a set of video image streams, including one that was driver-facing, which was used to collect glance behavior at 15 frames per second. All participants drove the same route in the greater Boston area. This route included several extensive segments of highway travel on I-93, where the data reported here was collected.

Baseline Driving & Secondary Tasks

Participants performed a variety of tasks as part of a larger assessment of in-vehicle interface use during highway driving [22]. This analysis specifically examined a voice dialing task sequence and subsequent story task. Prior to the start of this sequence, on the same segment of roadway, 2-minutes of single task driving ('just driving') was recorded. Participants then were asked to initiate a call using the voice interface, which connected them to a 25-s story message played from a remote voicemail system. After listening to this message, they were asked to leave a voice message telling back all they could remember of the story, and informed that what they said would be recorded and scored later. After ending the call, they were asked to reflect on the story they had just heard and to try to remember if they had forgotten to mention any details in their recorded message. They were asked not to say anything out loud and to simply think back carefully over all the details in the recording. The entire task sequence consisted of eight steps: pressing the hardware "push-to-talk" button located on the steering wheel, saying "phone", saying "call contact", saying "contact 4"/"contact 5", listening to the story, retelling the story, pressing the "end" button on the center stack touchscreen, and reflecting on the story.

Participants were asked to say either "contact 4" or "contact 5" based on a between-subjects manipulation of audio quality. Half of the participants heard a clear recording of the story message, and the other half heard a distorted recording of the message to mimic a poor phone connection (see supplemental material for audio tracks). In both audio quality conditions, the story participants heard was a single story

item selected from the Wechsler Logical Memory Task [36]. Its associated formal scoring system was used to examine the content of the recorded message. A score from 0-25 points was assigned based on the number of accurately recalled details.

Modified AttenD Algorithm

The AttenD algorithm takes a set of continuous glances over a task period and assigns a scalar value from 0.0 to 2.0 at the sample rate. This value changes based on a set of rules that depend on glance duration and location. Initialized at 2.0, for each second of off-road glance, the value decreases at a rate of one unit per second with a lower bound of 0.0. For return glances to the road, the value increases at a rate of one unit per second after a fixed latency period of 0.1s, which reflect a psychological refractory period from the cost of switching attention back to the roadway, with an upper bound of 2.0. If a driver glances at the mirrors or the speedometer, there is a latency of 1s before the scalar value decreases. By design, the AttenD algorithm is intended to predict distraction potential in which a 0.0 value demarks a distracted state [11]. However, in its ability to represent glance history in the form of a profile (or "buffer") that dynamically changes to reflect the amount of stored information about the roadway, it ties to other processes of attention and memory that are at play in how drivers sample information to form, retain, and update a robust representation of the driving environment [12,26]. The rule for the rate at which the scalar value increments when glance is returned on-road from off-road was modified from described above to reflect several experimentally-derived findings on the length of time required to regain situation awareness, as well as the rule for the latency of delay in buffer value based on location and duration of sampling to off-road regions. Finally, a rule was added to account for the distribution of glances to center and peripheral regions of the forward road to capture the effects of gaze concentration that result from changes in cognitive load [1,35,37].

Glance Coding

Glances were coded manually from the driver-facing video using a technique called dual coding with mediation [22], and used as input to the modified AttenD algorithm. Twelve glance regions were coded: (a) left road, (b) right road, (c) center stack, (d) instrument cluster, (e) rearview mirror, (f), right, (g), left, (h) right blind spot, (i) left blind spot, (j) passenger (the experimenter location in the left rear passenger seat), (k) other, and (l) eyes not visible.

Experimental Design

Independent Variables

To assess differences in cognitive load type, the story task was segmented into its three components for each participant: (1) the initial 25-s listening period, (2) the variable-length vocal response or "retelling" period (min = 13.0s; max = 88.5s; mean = 33.0s), and (3) the 60-s reflecting period. The 2-minute driving baseline that preceded the start

of the voice dialing sequence was included as a reference period. For reported analyses, a 2 x 3 mixed design was adopted: audio quality (clear; distorted) – a between-subjects factor; task type (listening; retelling; reflecting) – a within-subjects factor. Baseline driving values are provided for reference but were not considered in the primary analysis of cognitive effects.

Dependent Variables

The modified AttenD algorithm (or "cognitive buffer") was used to calculate buffer mean and standard deviation (SD) scores, and to produce aggregate profiles across participants per task period. Glances to select locations were analyzed during task periods. Measures of glance duration and frequency to central and peripheral regions of roadway are reported to assess effects of gaze centralization, which indicates an increase in cognitive load, in order to validate algorithm performance. The Wechsler Logical Memory Task scores were also analyzed.

Hypotheses

A higher cognitive load, reflected in a lower buffer mean and higher SD, is expected for retelling than for listening and reflecting periods of the story task, and for distorted over clear audio. An analysis of the distribution of glances between forward right and left regions of the driving scene is expected to indicate narrowed gaze dispersion for increased cognitive load as an underlying attentional element built into the modified buffer algorithm. Wechsler Logical Memory Task scores are expected to be lower for distorted compared to clear audio.

RESULTS

All analyses were performed using R. For the contrasts between tasks, twenty-four repeated measures t-tests were performed with Bonferroni corrections, which are reported in Table 1.

Cognitive Buffer Mean and SD

The AttenD buffer variant used here revealed differences between the three levels of cognitive load imposed by the tasks. Both the listening and reflecting tasks were associated with higher cognitive buffer scores (presumed lower cognitive load) than the retelling task for participants in both the clear audio and the distorted audio conditions. Similarly, both the listening and reflecting tasks were associated with lower standard deviation of cognitive buffer than the retelling tasks, again for both clear and distorted audio groups. No differences between clear and distorted audio were observed on the mean or standard deviation measures.

For the two-minute baseline drive periods, the average cognitive buffer score was 1.88 ($SE = 0.016$), and the average cognitive buffer standard deviation was 0.26 ($SE = 0.022$). In Figure 1, baseline task performance is indicated with a dashed line.

Road Left (Center) Glances

No significant differences were observed between the three tasks in terms of glances made to road center. However, we evaluated the effect of audio quality on mean single glance durations (MSGD) to road center using Welch's two sample t-tests, revealing that the clear audio group had significantly longer road center glances during listening, $t(27.09) = 2.90$, corrected $p = .044$.

Thus, while the cognitive buffer differentiated retelling from both listening and reflecting across both types of audio conditions, mean single on-road glance duration was better suited for distinguishing the distorted audio from clear audio conditions when its effects were presented (i.e., during the listening task).

Road Right Glances

To examine whether cognitive load influenced scan patterns, we also looked at glances to the right side of the windshield (i.e., road right) during the story task. Drivers spent significantly more time glancing to road right during the reflecting task than the retelling task. No differences between the audio conditions were observed.

Overall, these results suggest that the effect of story listening and story retelling on cognitive load can be seen through glance patterns to potentially situation-relevant windshield locations during auditory tasks, and that this effect is limited to tasks when audio is presented clearly. This also suggests that the effect of audio quality on gaze patterns is not reflected in glances to potentially situation-relevant locations.

Center Stack Glances

To further evaluate why on-road glance duration differed between some of the conditions, we examined the number of glances made to the center stack, where the visual prompts for the contact calling task were displayed. Because task length varied between tasks and between participants, the number of glances were normalized by task duration, producing a glance rate (per second). For the clear speech participants, the reflection task was associated with significantly fewer glances per second than the retelling task. For participants in the distorted audio group, both the listening and retelling tasks were associated with more glances to the center stack than the reflection task.

Contrasting the center stack glance rate between the clear and distorted audio groups, Welch's two sample t-tests revealed significant differences during the listening task, with distorted audio participants making more center stack glances than clear audio participants, $t(53.97) = 2.87$, corrected $p = .035$.

Aggregate Cognitive Buffer

The time course of the cognitive buffer was averaged across participants to visualize attention over time for the set of story tasks. Because the retelling task varied in length across participants, only the first 45s of the listening + retelling period were aggregated to generate a representative picture

of mean effects. For each time point in the plot, each participant's cognitive buffer score was averaged. Separate lines were generated per audio quality condition. The resulting profiles were smoothed using a polynomial function. As is evident in Figure 2, the buffer score dropped considerably in the first 5-10 seconds of both tasks before recovering. An analysis of video for each participant revealed that the initial drop in buffer score during the first 10 seconds of the listening portion is due to an unexpected visual-manual engagement – in which all 54 participants were observed to reach to the center stack to turn up the volume after the start of the story message.

Measure	Audio	Comparison	t	df	p (Uncorrected)	p (Bonferroni-corrected)	Mean Difference	Cohen's d
Mean Cognitive Buffer	**Clear**	**Listening vs. Retelling**	**3.91**	**24**	**0.00066**	**0.01584**	**0.35**	**0.78**
	Clear	Listening vs. Reflecting	0.01	24	0.99		0.0007	0.00
	Clear	**Retelling vs. Reflecting**	**4.99**	**24**	**0.00042**	**0.01008**	**0.35**	**1.00**
	Distorted	**Listening vs. Retelling**	**3.70**	**30**	**0.00087**	**0.02088**	**0.21**	**0.66**
	Distorted	Listening vs. Reflecting	2.35	30	0.026		0.17	0.42
	Distorted	**Retelling vs. Reflecting**	**5.45**	**30**	**0.0000065**	**0.000156**	**0.38**	**0.98**
SD Cognitive Buffer	**Clear**	**Listening vs. Retelling**	**4.12**	**24**	**0.00039**	**0.00936**	**0.21**	**0.82**
	Clear	Listening vs. Reflecting	0.21	24	0.84		0.015	0.04
	Clear	**Retelling vs. Reflecting**	**3.40**	**24**	**0.0023**	**0.0552**	**0.22**	**0.68**
	Distorted	**Listening vs. Retelling**	**4.34**	**30**	**0.00015**	**0.0036**	**0.17**	**0.78**
	Distorted	Listening vs. Reflecting	1.16	30	0.25		0.073	0.21
	Distorted	**Retelling vs. Reflecting**	**5.13**	**30**	**0.000016**	**0.000384**	**0.24**	**0.92**
Center Stack Glances per Second	Clear	Listening vs. Retelling	0.02	24	0.98		0.00037	0.00
	Clear	Listening vs. Reflecting	2.72	24	0.012		0.047	0.54
	Clear	Retelling vs. Reflecting	3.13	24	0.0046		0.047	0.63
	Distorted	Listening vs. Retelling	2.97	30	0.0058		0.038	0.53
	Distorted	**Listening vs. Reflecting**	**7.74**	**30**	**1.245E-08**	**2.988E-07**	**0.11**	**1.39**
	Distorted	**Retelling vs. Reflecting**	**6.54**	**30**	**3.151E-07**	**7.5624E-06**	**0.07**	**1.17**
Road Right Glance Time per Second	Clear	Listening vs. Retelling	2.08	24	0.049		0.013	0.42
	Clear	Listening vs. Reflecting	2.75	24	0.011		0.038	0.55
	Clear	**Retelling vs. Reflecting**	**4.14**	**24**	**0.00037**	**0.00888**	**0.51**	**0.83**
	Distorted	Listening vs. Retelling	0.50	30	0.62		0.0082	0.09
	Distorted	Listening vs. Reflecting	1.24	30	0.22		0.16	0.22
	Distorted	Retelling vs. Reflecting	0.32	30	0.75		0.008	0.06

Table 1. Glance metrics by audio condition and task. Bonferroni corrections were applied to ensure a family-wise Type I error-rate of 0.05 across the study; p values were multiplied by 24 (the number of total comparisons).

Downstream from this initial drop, for most of the listening task, buffer scores are far higher for the clear speech participants. During the retelling task, buffer scores for both groups drop at similar rates, although the distorted speech participants show a lower minimum score, on average. Figure 2 visually depicts the process of the buffer stabilizing during the cognitive load associated with listening, with a subsequent and continued decline once participants engage in story retelling. During the reflecting period, the buffer remains stable at a higher score.

Wechsler Logical Memory Task Scores
An analysis of Wechsler Logical Memory Task scores showed that drivers presented with the clear audio during the story message had significantly higher scores, reflecting their

more accurate recall of story details than those participants who heard the distorted audio (clear audio quality: $M = 13.0$, $SE = 0.66$; distorted audio quality: $M = 4.42$, $SE = 0.79$; $t(54) = 8.33$, $p < .0001$). Distorted audio during the listening period of the story task interfered with participants' ability to

accurately store and retain key details. These results point to an increase in task difficulty when audio quality is poor.
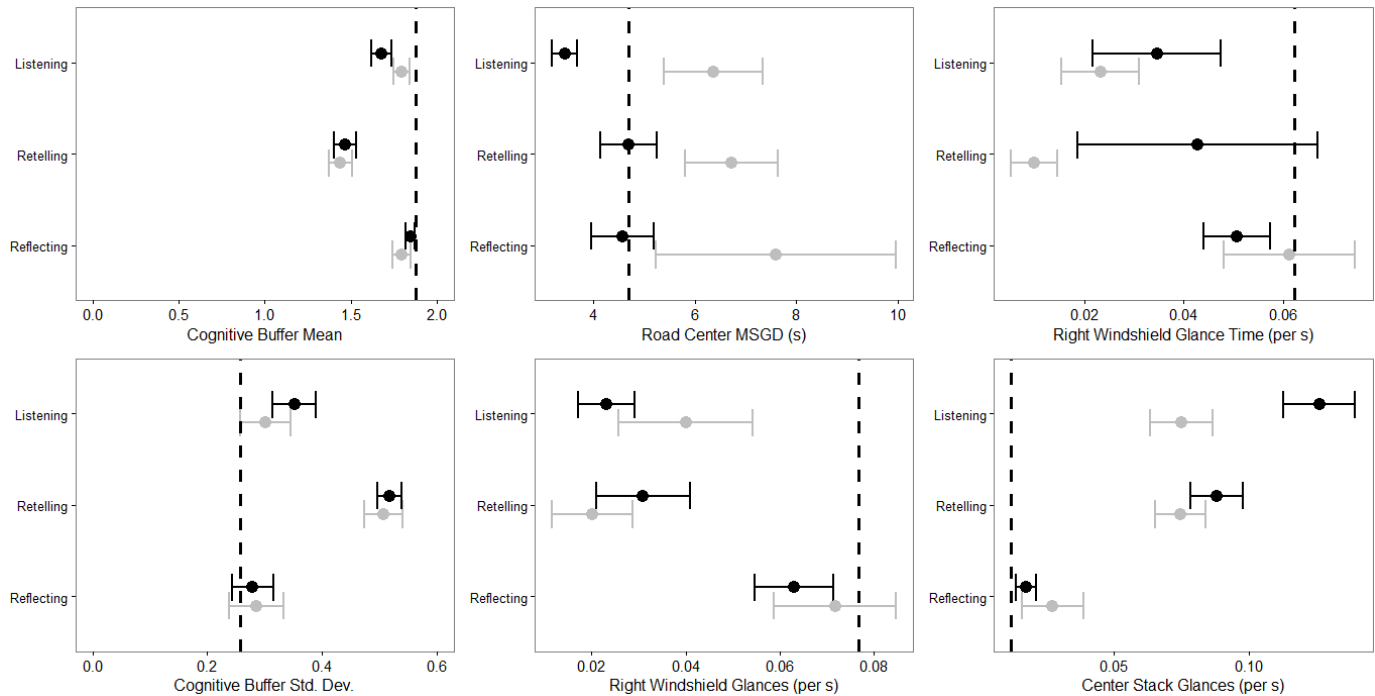
Figure 1. Glance metrics by task and audio quality. Grey points represent clear audio participants, and black points represent distorted audio participants. Error bars represent standard error. Dashed lines represent baseline task performance.

DISCUSSION

The modified AttenD buffer variant used here proved sensitive to different levels of cognitive load – and may be useful to HMI practitioners who are working to develop and evaluate HMIs that will support drivers in managing their attention to the road, and in the development of real-time driver attention monitoring systems. Cognitive buffer mean and SD scores indicated increased load for retelling versus listening and reflecting portions of the story task, consistent with our hypotheses and in keeping with previous findings that speaking tasks are more cognitively loading than listening or mental reflection tasks while driving [3,9,25,30]. While this distinction did not show statistical significance for the metrics included to test for effects of gaze centralization for drivers' access to central and peripheral regions of the road (in particular, for the listening vs. retelling comparison), the non-corrected p-values in Table 1 indicate the presence of this effect and its underlying role in the observed buffer metric differences. In this work, it can also be seen that for some levels of cognitive load (e.g., reflecting on a story, or listening to clear audio of a story), there is little to no effect on buffer scores relative to baseline driving – whereas for the speech production task (like retelling a story), there is a

larger and significant drop in buffer score from baseline (to a level between 1.4 - 1.5), and an increase in variability of buffer scores. A buffer at this level (during a demanding speech production task like story retelling) would indicate that drivers may have less than ideal amounts of situation information stored in memory, or in their representations of the driving situation, perhaps leaving them vulnerable to making incorrect forecasts about where to look in the next moments of driving, or what to expect in the unfolding driving scenario [26].

In interpreting these findings, it is important to consider the effects of cognitive load in the context of other types of task loading. For example, how does the story reflection task compare to the type of cognitive activity involved in "thinking," or "lost-in-thought" or "mind-wandering" that may sometimes occur during "just driving"? If so, how does it relate to crash risk?

It is also important to question how these effects of cognitive load compare to the effects of visual-manual load. Zhang et al. [38] examined visual-manual task loadings using the AttenD buffer. Specifically, they evaluated five matched visual-manual task types for each of two different HMI

implementations (a rotary haptic prototype HMI and a touchscreen HMI). These visual-manual tasks led to a range of effects on driver attention – but in general produced much larger drops in buffer score than any of the cognitive tasks examined here in this study. For example, mean median buffer scores for the visual-manual tasks ranged from approximately 0.60 to 0.73 across the two types of visual-manual HMIs in the Zhang et al. [38] work. Most importantly, however, the AttenD algorithm was sensitive to task type variation among the visual-manual task types and

between the visual-manual interfaces – and provided insights into HMI improvements that could be made in task structure and HMI element design toward supporting driver attention. The application of the AttenD algorithm in this study shows similar diagnostic value for detecting the presence of cognitive load in driver interaction with HMIs. In particular, the modified AttenD approach emerged as particularly useful in determining what levels of cognitive load influence attention to the road, and how.
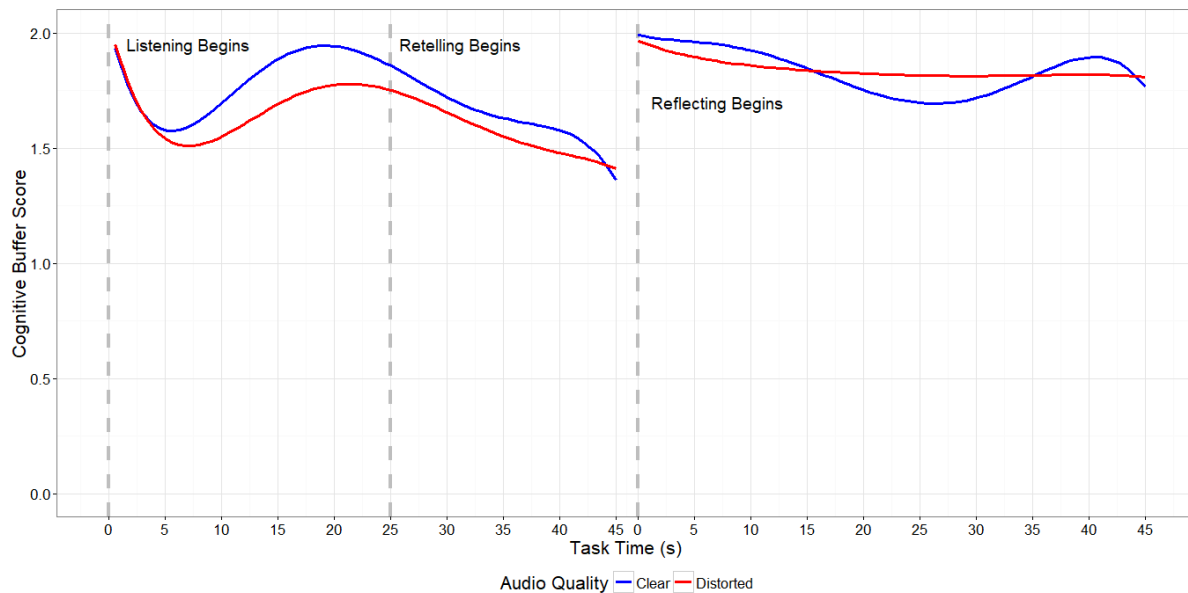
Figure 2. Aggregate cognitive buffer for listening, retelling, and reflecting periods of the story task. Individual buffer profiles were averaged at each second of time to generate an aggregate profile per audio quality condition.

CONCLUSION

While this work indicates a potential value in the modified AttenD buffer for HMI evaluation and driver monitoring applications, further exploration of the approach using a wider and potentially more structured range of cognitive tasks is needed. Work reported on elsewhere [12,26] shows that the buffer has a conceptual link to safety (in settings where naturalistic data of accidents exist). Greater quantities of naturalistic accident data with moment-to-moment driver glance profiles that encompass periods of known cognitive demand may eventually allow for further consideration of the relationship between attention allocation and safety using the buffer metrics. Ongoing work is currently investigating a range of factors central to the application of the buffer in device design and real-time attention management. This multifaceted effort aims to further optimize the sensitivity of the buffer to changes in situation awareness. Efforts are also underway to consider the impact of environmental characteristics, vehicle automation levels, and inter-related other factors on driver attention using the approaches discussed here. In these efforts, the buffer is being developed

as an attention management approach that is responsive to both the high and low portions of the demand curve for traditionally manually-controlled vehicles and ones equipped with automated features that require driver involvement. This focus allows us to move the language of demand assessment from one focused on distraction to one that emphasizes driver attention management and safe operation. The overarching aim is to consider, within the context of a single safety-centric model, the demands placed on drivers, active safety systems, other higher order forms of automation, and systems designed to support driver state.

ACKNOWLEDGMENTS

REFERENCES

1. Angell, L., Perez, M. A., & Soccolich, S. A. (2015). Identification of Cognitive Load in Naturalistic Driving. *NSTSCE; 15-UT-037*.

2. Caird, J. K., Willness, C. R., Steel, P., & Scialfa, C. (2008). A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis & Prevention*, *40*(4), 1282-1293.

3. Chang, C.-C., Sodnik, J., & Boyle, L.N. (2016). Don't speak and driver: Cognitive workload of in-vehicle speech interactions. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 99-104.

4. Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. In *Proceedings of the National Academy of Sciences*, *113*, 2636–2641.

5. Engstrom, J., Markkula, G., Victor, T., & Merat, N. (2017). Effects of cognitive load on driving performance: The cognitive control hypothesis. *Human Factors, 59*(5), 734-764.

6. Harbluk, J. L., Noy, Y. I., & Eizenman, M. (2002). *The impact of cognitive distraction on driver visual behaviour and vehicle control*. Technical report, Transport Canada, Ottawa.

7. He, J., Bebic, E., Lee, Y.-C., & McCarley, J. S. (2011). Mind wandering behind the wheel: Performance and oculomotor correlates. *Human Factors*, *53*, 21–31.

8. Hickman, J. S., Hanowski, R. J., & Bocanegra, J. (2010). *Distraction in commercial trucks and buses: Assessing prevalence and risk in conjunction with crashes and near-crashes* (Report No. FMCSA-RRR-10-049). Washington, DC: Federal Motor Carrier Administration.

9. Hsieh, L., Young, R. A., Bowyer, S. M., Moran, J. E., Genik, R. J., Green, C. C., & Seaman, S. (2009). Conversation effects on neural mechanisms underlying reaction time to visual events while viewing a driving scene: fMRI analysis and asynchrony model. *Brain Research*, *1251*, 162-175.

10. Kidd, D. G., & McCartt, A. T. (2015). The relevance of crash type and severity when estimating crash risk using the SHRP2 naturalistic driving data. In *Proceedings of the 4th International Driver Distraction and Inattention Conference*, Sydney, Australia.

11. Kircher, K., & Ahlström, C. (2013). *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.

12. Lee, J., Sawyer, B. D., Mehler, B., Angell, L., Seppelt, B. D., Seaman, S., & Reimer, B. (2017). *Linking the Detection Response Task and the AttenD Algorithm through the Assessment of Human-Machine Interface Workload* (No. 17-06664).

13. Mazzae, E. N., Goodman, M., Garrott, W. R., & Ranney, T. A. (2005). NHTSA's research program on wireless phone driver interface effects (Paper No. 05-0375). In *Proceedings of the 19th International Technical Conference on the Enhanced Safety of Vehicles* (CD-ROM). Washington, DC: National Highway Traffic Safety Administration.

14. McCarley, J. S., Vais, M. J., Pringle, H., Kramer, A. F., Irwin, D. E., & Strayer, D. L. (2004). Conversation disrupts change detection in complex traffic scenes. *Human Factors*, *46*(3), 424-436.

15. Mehler, B., Kidd, D., Reimer, B., Reagan, I., Dobres, J. & McCartt, A. (2016). Multi-modal assessment of on-road demand of voice and manual phone calling and voice navigation entry across two embedded vehicle systems. *Ergonomics, 59*(3), 344-367. doi:10.1080/00140139.2015.1081412.

16. Muttart, J. W., Fisher, D. L., Knodler, M., & Pollatsek, A. (2007). Driving without a clue: Evaluation of driver simulator performance during hands-free cell phone operation in a work zone. *Transportation Research Record: Journal of the Transportation Research Board*, *2018*, 9–14.

17. National Highway Traffic Safety Administration. (2013). *Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices*. Publication NHTSA-2010-0053. National Highway Traffic Safety Administration, Washington, D.C.

18. Olson, R. L, Hanowski, R. J., Hickman, J. S., & Bocanegra, J. (2009). *Driver distraction in commercial vehicle operations* (Report No. FMCSA-RRR-09-042). Washington, DC: Department of Transportation, Federal Motor Carrier Safety Administration.

19. Ranney, T. A., Baldwin, G. H. S. , Smith, L. A., Mazzae, E. N., & Pierce, R. S. *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.

20. Ranney, T. A., Harbluk, J. L., & Noy, Y. I. (2005). Effects of voice technology on test track driving performance: Implications for driver distraction. *Human Factors*, *47*(2), 439-454.

21. Reimer, B., Angell, L., Strayer, D., Tijerina, L. & Mehler, B. (2016). Evaluating Demands Associated with the Use of Voice-Based In-Vehicle Interfaces. In *Proceedings of the 60th Annual Meeting of the Human Factors and Ergonomics Society.* Washington, DC. pp. 2083-2087.

22. Reimer, B., Mehler, B., Dobres, J., & Coughlin, J.F. (2013). *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.

23. Reimer, B., Mehler, B., Dobres, J., McAnulty, H., Mehler, A., Munger, D., & Rumpold, A. (2014).

Effects of an "expert mode" voice command system on task performance, glance behavior & driver physiology. In *Proceedings at the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, Seattle, WA.

24. Reimer, B., Mehler, B., Reagan, I., Kidd, D., & Dobres, J. (2016). Multi-modal demands of a smartphone used to place calls and enter addresses during highway driving relative to two embedded systems. *Ergonomics*, *59*(12), 1565-1585.

25. Salvucci, D. D., & Beltowska, J. (2008). Effects of memory rehearsal on driver performance: Experiment and theoretical account. *Human Factors*, *50*, 834–844.

26. Seppelt, B.D., Seaman, S., Lee, J., Angell, L., Mehler, B., & Reimer, B. (2017). Glass half-full: On-road glance metrics differentiate crashes from near-crashes in the 100-car data. *Accident Analysis & Prevention, 107*, 48-62.

27. Shinar, D., Tractinsky, N., & Compton, R. (2005). Effects of practice, age, and task demands, on interference from a phone task while driving. *Accident Analysis & Prevention*, *37*(2), 315-326.

28. Sodhi, M.S., Reimer, B., & Llamazares, I. (2002). Glance Analysis of Driver Eye Movements to Evaluate Distraction. *Behavior Research Methods, Instruments and Computing,* 34(4), 529-538.

29. Stanley, L. M., Kelly, M. J., & Lassacher, S. (2005). Driver performance while interacting with the 511 travel information system in urban and rural traffic. In *Proceedings of the Third International Driving Symposium on Human Factors in Driving Assessment, Training and Vehicle Design* (pp. 486-492).

30. Strayer, D.L., Biondi, F., & Cooper, J.M. (2017). Dynamic workload fluctuations in driver/non-driver conversational dyads. In *Proceedings of the Ninth International Driving Symposium on Human Factors in Driving Assessment, Training and Vehicle Design*.

31. Strayer, D. L., Drews, F. A., Crouch, D. J., Johnston, W. A., Walker, T. A. I. W., & Herrmann, D. (2005). Why do cell phone conversations interfere with driving? *Cognitive technology: Essays on the transformation of thought and society*, 51-68.

32. Strayer, D. L., Drews, F. A., & Johnston, W. A. (2003). Cell phone-induced failures of visual attention during simulated driving. *Journal of Experimental Psychology: Applied*, *9*, 23–52.

33. Victor, T. (2005). *Keeping eye and mind on the road*. Retrieved from Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences 9.

34. Victor, T., Bärgman, J., Boda, C-N., Dozza, M., Engström, J., Flanagan, C., Lee, J. D., & Markkula, G. (2015). *Analysis of Naturalistic Driving Study data: Safer glances, driver inattention, and crash risk* (SHRP 2 Safety Project S08A). Washington, DC: Transportation Research Board.

35. Victor, T., Harbluk, J., & Engström, J. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F*, *8*, 167–190.

36. Wechsler, D., (1997). *Wechsler Adult Intelligence Scale III [Manual]. 3rd ed.*, San Antonio, TX: The Psychological Corporation.

37. Wang, Y., Reimer, B., Dobres, J. & Mehler, B. (2014). The sensitivity of different methodologies for characterizing drivers' gaze concentration under increased cognitive demand. *Transportation Research Part F: Traffic Psychology and Behaviour,* 26 (Part A), pp. 227-237.

38. Zhang, Y., Angell, L., Pala, S., Hara, T. et al. (2017). Can You Still Look Up? Remote Rotary Controller vs. Touchscreen. *SAE Technical Paper 2017-01-1386*.