

Jun 29th, 12:00 AM

It's All in the Timing: Using the Attend Algorithm to Assess Texting in the Nest Naturalistic Driving Database

Sean Seaman
Touchstone Evaluations, Inc.

Joonbum Lee
MIT AgeLab and New England University Transportation Center

Bobbie Seppelt
Touchstone Evaluations, Inc.

Linda Angell
Touchstone Evaluations, Inc.

Bruce Mehler
MIT AgeLab and New England University Transportation Center

See next page for additional authors

Follow this and additional works at: <http://ir.uiowa.edu/drivingassessment>

Seaman, Sean; Lee, Joonbum; Seppelt, Bobbie; Angell, Linda; Mehler, Bruce; and Reimer, Bryan. It's All in the Timing: Using the Attend Algorithm to Assess Texting in the Nest Naturalistic Driving Database. In: Proceedings of the Ninth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, June 26-29, 2017, Manchester Village, Vermont. Iowa City, IA: Public Policy Center, University of Iowa, 2017: 403-409. <https://doi.org/10.17077/drivingassessment.1665>

This Event is brought to you for free and open access by the Public Policy Center at Iowa Research Online. It has been accepted for inclusion in Driving Assessment Conference by an authorized administrator of Iowa Research Online. For more information, please contact lib-ir@uiowa.edu.

Presenter Information

Sean Seaman, Joonbum Lee, Bobbie Seppelt, Linda Angell, Bruce Mehler, and Bryan Reimer

IT'S ALL IN THE TIMING: USING THE ATTEND ALGORITHM TO ASSESS TEXTING IN THE NEST NATURALISTIC DRIVING DATABASE

Sean Seaman¹, Joonbum Lee², Bobbie Seppelt¹, Linda Angell¹, Bruce Mehler², & Bryan Reimer²

¹Touchstone Evaluations, Inc.

²MIT AgeLab and New England University Transportation Center

Summary: To better understand cellular phone texting behavior and its relationship to crashing, we combined the sample-level glance data of NEST with the AttenD buffer algorithm to visualize glancing during texting within naturalistic epochs ending in crashes or no crashes. We found that texting periods were quite similar across the two, both in duration, number of individual texting tasks, and overall shape of the AttenD buffer curve. However, we found that crash epoch texting tended to occur closer to the onset of a crash event, and that texting during crashing may be initiated when the AttenD buffer level is lower (indicating depleted situation awareness), possibly due to prior or ongoing operational or secondary activities. We also made similar comparisons for radio interaction tasks, and found substantial differences between radio crash and baseline interactions. We conclude that whether a texting period ends in a crash may be dependent upon more than the individual differences in length of texting or amount of glancing. One's level of situation awareness at the start of the activity (indicating a potential lack of judgment in picking up the device), in combination with a cascading losses of situation awareness that arise from the temporal pattern of on-road and off-road glances upstream from a safety-critical event, may be key predictive factors.

OBJECTIVES

Naturalistic databases, such as 100 Car and SHRP2, provide a unique opportunity to examine driving behavior in non-contrived settings; however, too often those data, originally recorded with rich video and laboriously coded by observers, are aggregated to single data points. Using the Naturalistic Engagement in Secondary Task (NEST) database (Owens et al., 2015), a subset of the Strategic Highway Research Program (SHRP2) database focused primarily on secondary task engagement during crashes and near-crashes, we explored driver glancing behavior in one of the more commonly observed secondary tasks—cellular phone texting—and compared this to glancing behavior during radio interactions. To better visualize how attention is distributed over time, we sought out an algorithm that computes an ongoing score over time based on on-road and off-road glances: the AttenD algorithm developed by Kircher and Ahlström (2009).

Prior work (e.g., Victor et al., 2014; Liang, et al., 2012; Yekhshatyan, 2010), has suggested that what goes wrong in some distraction-related crashes is that a glance away from the road occurs at an inopportune time (e.g., when a precipitating event occurs and/or situation kinematics change very rapidly). However, if this is a characteristic of what goes wrong when a secondary activity ends in a crash, what leads drivers to look away at that inopportune moment? Does the pattern of glances upstream from a precipitating event lead to a possible loss of situation awareness, and does a loss of situation awareness result in cascading effects, which lead to poor

timing and placement of glances over time? These are the types of questions that can be explored in the context of a buffer model like the AttenD algorithm that we used here.

The AttenD algorithm was selected for several reasons. The algorithm yields a hybrid metric that reflects both off-road and on-road glancing, and recent research (Seppelt et al., 2017) suggests that patterns in *on-road* glancing are important at differentiating safety critical events in naturalistic data. As such, the AttenD algorithm uses a buffer concept to represent the information that a driver can encode from the driving situation during “on-road” glances and the loss of information when the driver looks away from the road. This buffer value has been tied by Seppelt et al. (2017) as a measure indicative of situation awareness. The algorithm is one of the few proposed to date that places an emphasis on how the *temporal pattern of a driver’s glances* affects the amount and quality of information that is available to the driver at any moment. Since this type of information may play a role in where and when a driver chooses to glance next, the algorithm was a good match for addressing the questions of interest here. Using this algorithm, we visualized texting and radio interactions during Baseline and Crash epochs in NEST, and used this to better understand what goes wrong when texting ends in a crash event.

METHODS

The NEST database contains a subset of the SHRP2 safety-critical event (SCE) data, along with four Baseline epochs (i.e., epochs not linked to crashes or near-crashes) from each driver for each of that driver’s independent observations in the SCE (Crash and Near-Crash epoch) set. This subset was specifically compiled to provide SCE and Baseline data for drivers who had distraction-related SCEs; thus, all the SCE epochs contain secondary task activity (e.g., manual tasks, auditory/vocal tasks). Baselines epochs were of two types: Some Baseline epochs contained secondary task activity, while some did not. While a subset of SHRP2, NEST contains a greater level of detail for a longer period about its selected epochs than SHRP2 overall. NEST epochs contain, typically, twenty seconds of glance and secondary-task coded data; for SCE epochs, this is typically extended to thirty seconds, although the additional ten seconds were coded after the onset of the precipitating event linked to the SCE. For the present study, all glances and secondary task engagements occurring after the precipitating event were discarded, as the precipitating event may itself have altered subsequent glance patterns. Baseline epochs were only 20 seconds long to begin with (since they did not contain a precipitating event or ensuing 10-second period), and thus were matched in length to the 20-second long SCE epochs.

The NEST database contains a combination of epoch- and sample-level data. At the epoch-level, NEST contains descriptions of Crash, Near-Crash, and Baseline epochs, including secondary task activities observed. At the sample-level (provided at 10 Hz), NEST also contains glance behavior (coded in terms of the location to which gaze was directed at each frame) and secondary task engagement (coded in terms of type of task or activity underway). For the present study, the focus is on Crash epochs coded as containing either “Texting on a cell phone” or “Adjusting/monitoring radio” activity. Two types of Baseline epochs were analyzed: those containing “Texting” or “Adjusting/monitoring radio” and those containing “No secondary task.” To ensure that the analysis utilized independent sets of epochs, Baseline epochs contributed by any of the same drivers appearing in the texting or radio Crash epochs were removed. (A fully-crossed comparison would have yielded too few epochs.) Similarly, Baseline epochs were

removed from the “No secondary task” set if they were contributed by drivers who also appeared in the Baseline set of texting epochs (or radio epochs, analyzed separately), to ensure all three groups were composed of independent observations. For all computations reported, drivers with multiple epochs in a condition were reduced to a single data point by averaging. For texting, this yielded 22 Crash epochs with texting across 22 drivers; 69 Baseline epochs with texting across 46 drivers; and 171 Baseline epochs with no tasks across 97 drivers. For radio interactions, this yielded 13 Crash epochs with radio interaction across 13 drivers; 59 Baseline epochs with radio interaction across 52 drivers; and 166 Baseline epochs with no tasks across 92 drivers. Sample-level data was arranged as per Seaman et al. (2016).

The AttenD algorithm was applied to this glance data on an epoch-by-epoch basis. For each epoch, the initial buffer value was set at 2, and for each second of off-road glances (glances not to the forward, right windshield, or left windshield areas of interest) the buffer value was decremented by 1 point, or 0.1 point per 10 Hz sample of off-road glancing. If the buffer value reached 0, it did not drop further until the driver glanced back to the windshield; then, it began increasing again, after a latency period of 0.1 seconds at a rate of 1 point per second, until it returned to 2 points. Mirror and instrument cluster glances yielded a delay in buffer depletion of 1 second, reflecting that these are regions that are important to situationally-aware driving. The application of this algorithm yielded an array of values representing attention for every sample of glance data in the Baseline and Crash epochs. Visualizations of the buffer data were made by truncating available data to twenty seconds and averaging each data point within those twenty seconds (200 samples). Means and standard errors were computed across epochs.

RESULTS

Table 1 contains statistics about the sub-epoch task data for texting and radio interactions. Task duration and the average number of texting tasks initiated within an epoch were roughly equivalent between Crash and Baseline epochs, which shows some consistency across texting from two different groups of drivers. For radio tasks, however, the difference was quite pronounced, with the average radio interaction lasting more than twice as long in Crash epochs than in Baseline epochs, suggesting that different radio tasks were being done during Baseline than Crash epochs (e.g., Perez, Angell, & Hankey, 2015).

Table 1. Sub-epoch task statistics

Condition	Average task duration	Average number of tasks	Average latest task end
Baseline – Texting	9.93 s (<i>sd</i> = 6.639 s)	1.11 (<i>sd</i> = 0.297)	3.578 s (<i>sd</i> = 5.807 s)
Crash – Texting	10.31 s (<i>sd</i> = 6.555 s)	1.14 (<i>sd</i> = 0.351)	1.266 s (<i>sd</i> = 3.229 s)
Baseline – Radio	3.42 s (<i>sd</i> = 3.779 s)	1.05 (<i>sd</i> = 0.210)	9.071 s (<i>sd</i> = 6.936 s)
Crash – Radio	7.38 s (<i>sd</i> = 7.006 s)	1.15 (<i>sd</i> = 0.376)	6.829 s (<i>sd</i> = 6.319 s)

The temporal alignment of the end of the last texting task differentiates the two types of epochs. In identifying task boundaries, the sample-level coding of task performance was used; when a

task continued until the end of an epoch (or beyond the precipitating event of a Crash epoch), it was coded as ending at the end of the available epoch. Overall, it appears Crash epoch texting tasks end closer to the end of an epoch (in this table, lower values indicate a closer ending to the onset of the precipitating event). This suggests a coordination of glancing across the Crash texting epochs that could yield a loss of situation awareness at the end of the Crash epochs, immediately before the precipitating event associated with crashing. Radio tasks, on the other hand, tended to end earlier in an epoch, with Crash radio epochs ending later.

Overall, mean AttenD buffer values across an epoch varied significantly with epoch condition, Baseline texting vs. Baseline no task vs. Crash texting, $F(2,162) = 94.78, p < .00001$; Baseline radio vs. Baseline no task vs. Crash radio, $F(2,154) = 37.43, p < .00001$, suggesting that the buffer is sensitive to task and safety-critical event aspects of an epoch. Table 2 has the mean buffer values and standard errors for each condition.

Table 2. Mean AttenD buffer values by epoch type and secondary task

Epoch Type	Texting - Mean AttenD Buffer Value	Texting - SEM	Radio - Mean AttenD Buffer Value	Radio -SEM
Crash	1.081	0.1345	1.420	0.1497
Baseline	1.234	0.0805	1.770	0.0401
Baseline (no task)	1.975	0.0072	1.973	0.0090

While both the texting and radio conditions differed significantly from the no secondary task Baseline condition ($p < .00001$ for all four comparisons), the two texting conditions did not differ significantly from one another, $F(1,66) = 1.097, p = 0.299$; the two radio conditions (Baseline vs. Crash) *did* differ significantly from one another, $F(1,63) = 1.276, p = 0.002$. Although the texting epochs lined-up in expected ways, with texting Crash epochs having lower AttenD buffer values than both Baseline conditions, only the effect of the texting task was significant on mean overall AttenD buffer value. This may partly be due to the small N (only 22 crashes) and variability in naturalistic driving epochs, but overall our interest was not in reducing the AttenD to a single timepoint for statistical comparison, but to visualize glance performance over time to identify trends in the AttenD buffer state across the three types of conditions, which we present in Figure 1. Note that all AttenD buffer lines begin at 2.0, the starting value of the AttenD buffer within an epoch, and decline from there over time as participants make glances off-road. For both Baseline no task conditions, this line asymptotes quickly and remains stable at a high value, suggesting that the starting value of 2 is appropriate under the (admittedly best-case) assumption that task activity is preceded by periods of no task activity. For the two texting conditions, this line declines much further. For the Crash texting epochs, it continues to decline throughout the epoch until the onset of the precipitating event of the crash. This provides a nice visual illustration of overall ongoing higher levels of glance disruption during texting, while the differences in texting conditions are subtler. The Crash epoch drivers appear to suffer from an onset of ill-timed late epoch texting that may have led to missing cues associated with the precipitating event. AttenD values drop for the two radio sets of epochs, but drop much further during Crash epochs (which is likely responsible for the main effect of Crash vs. Baseline for

radio). Again, Crash epoch drivers appear to suffer from the onset of ill-timed late epoch radio interaction that is likely associated with missing important cues. To investigate this possibility, we examined the sample-level secondary task coding to identify when drivers engaged in secondary tasks, and to isolate these periods for visualization with the AttendD buffer.

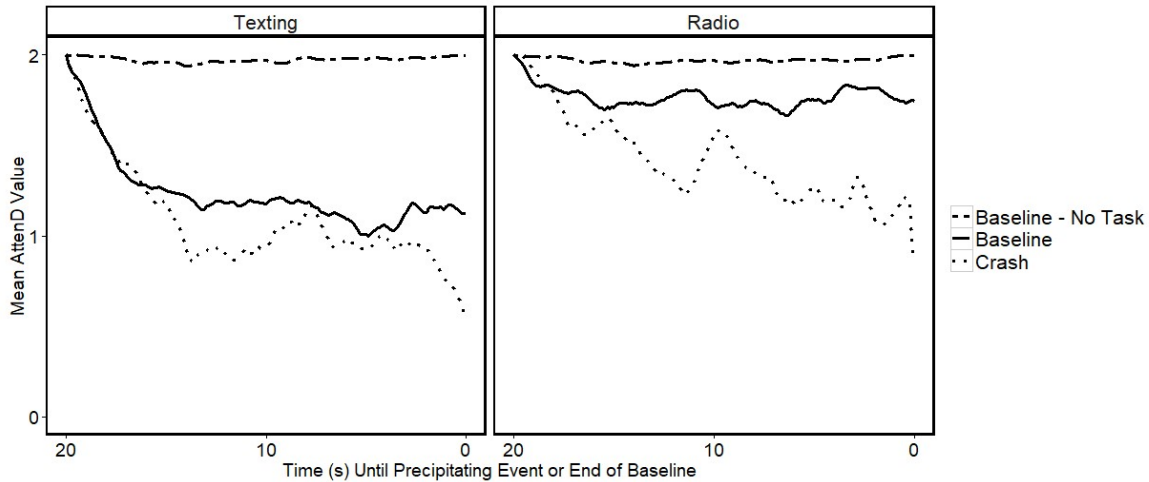


Figure 1. AttendD buffer value over time for epochs containing texting (left) and radio interaction (right)

Because tasks can have varying durations, we normalized each epoch’s AttendD buffer curve by first smoothing the data to reduce the influence of outliers (using the “loess” local polynomial regression fitting function in R (R Core Team, 2016), and then by using linear interpolation to normalize each varying length sample to 250 samples. These can be seen in Figure 2. Both Crash and Baseline epochs show texting AttendD curves of similar shapes, with steep declines early. Critically, the Crash AttendD curve starts at a lower buffer level ($\bar{x} = 1.45$) versus the Baseline curve ($\bar{x} = 1.70$). This roughly corresponds to the difference at task completion (0.37 and 0.61 for Crash and Baseline).

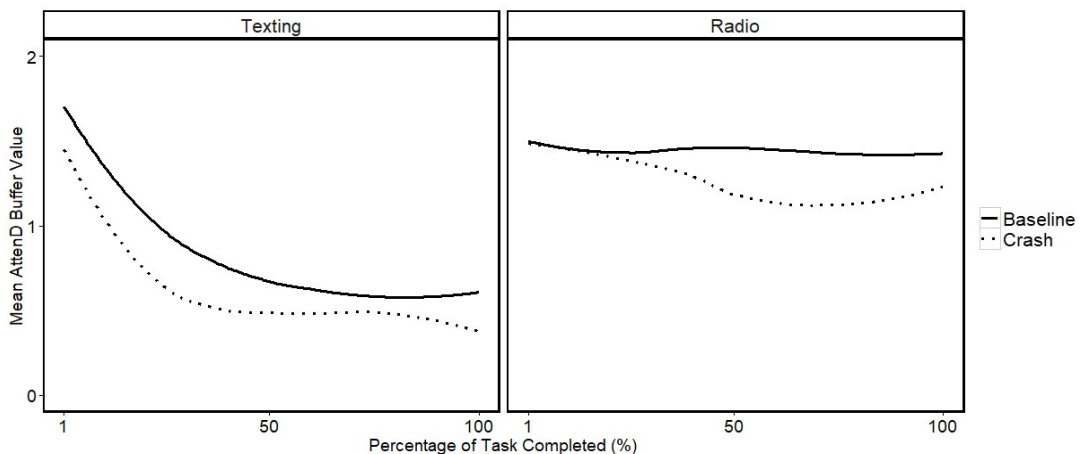


Figure 2. AttendD buffer values for texting (left) and radio (right) tasks in crash and baseline epochs

Thus, it appears that differences in texting between Crash and Baseline epochs are a difference in timing: a difference in *when* a task is initiated, relative to a SCE, and relative to *other* pre-task

activities. For those epochs that end in crash, this analysis reveals that the buffer value is at a lower value *before* a texting task is even initiated. For radio interactions, this trend is quite different, with Baseline radio interactions showing no noticeable loss in buffer value, and Crash radio interactions showing a marked decline; both Crash and Baseline radio interactions start at similar buffer values ($\bar{x} = 1.49$ and $\bar{x} = 1.50$ for Crash and Baseline interactions), but end at different values ($\bar{x} = 1.23$ and $\bar{x} = 1.43$ for Crash and Baseline interactions), suggesting that it is the different nature of radio interactions across the two sets of epochs, not the conditions of task onset.

To evaluate an explanation as to why Crash texting periods would start with a lower AttenD buffer value, given approximately the same number of texting tasks are initiated during Crash and Baseline epochs, we looked at *other* tasks that are performed within an epoch (such as talking), potentially simultaneously with texting, and compared task counts. Crash texting epochs were associated with more comorbid tasks ($\bar{x} = 2.53$, $sd = 1.125$) than Baseline epochs ($\bar{x} = 2.09$, $sd = 1.042$), although this difference was not significant, $F(1,59) = 2.037$, $p = 0.159$.

CONCLUSION

The AttenD algorithm successfully differentiated texting from no task epochs, whether that texting occurred in a Crash or Baseline epoch. Furthermore, plotting epochs and tasks using the AttenD algorithm provided insight into why texting sometimes ends in crash: namely, the lower initial level of situation awareness and the corresponding loss of information about the driving situation, as illustrated by the AttenD profiles during texting. During crashes, the low-level situational awareness most likely happens to coincide with an event on the roadway that full attention to the roadway may have identified. Texting, whether before a Crash or during a Baseline period, appears to be uniform: It is essentially the timing of the activity that may determine how badly it ends. “Timing” in this context means more than it is often taken to mean. The AttenD buffer analyses revealed that: one, for epochs which ended in crash, texting started at a lower buffer value, indicating that the driver’s awareness of the situation was already poorer than in Baseline, and the choice to initiate the task may reflect this; and two, for these epochs, begun at a lower level of situation awareness, the basis on which the driver will calculate when and where to look next (and for how long) may also be compromised. This may lead to increasingly poor choices over the course of the task about each ensuing glance, and ultimately lead to off-road glancing at an inopportune time when the situation changes, perhaps rapidly.

One insight this analysis revealed was that texting during crashes is started at a lower level of situation awareness than texting during baseline periods. Initial explorations suggest that this may be due to the driver’s engagement in *other* comorbid secondary tasks before texting is initiated during epochs that end in crashes. While it’s possible that the extra half a task per epoch could be responsible for the difference in initial AttenD buffer value, which could be responsible for the poor choice of timing that leads to Crash texting to occur close to precipitating events of crashes, a lack of statistical significance at this point only leaves this a possibility to be explored by examining more secondary task activity. However, it is an important one to explore with future work on how individual differences (e.g., driving inexperience, young drivers’ ability to scan for and recognize hazards) may enter into choices about when to engage in secondary tasks.

While the AttenD buffer provided insight into epoch- and task-level activity in the NEST dataset, our efforts are currently focused on improving the algorithm to identify periods of problematic situation awareness and applying it to HMI evaluation and real-time driver state estimation. Using naturalistic data, where the ground truth of the outcome of task engagement is known, we can ensure that algorithm has safety efficacy. Ongoing work is integrating measures of cognition into the model as a more comprehensive measure of attention and refining our understanding of the enhanced algorithm's performance across many secondary tasks to identify how situation awareness changes across the driving task in cases where something goes right and wrong.

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

- Liang, Y., Lee, J. D., & Yekhshatyan, L. (2012). How dangerous is looking away from the road? Algorithms predict crash risk from glance patterns in naturalistic driving. *Human Factors*, 54, 1104-1116.
- Kircher, K., & Ahlström, C. (2009). Issues related to the driver distraction detection algorithm AttenD. In *1st Int. Conf. on Driver Distraction & Inattention*. Gothenburg, Sweden.
- Owens, J. M., Angell, L., Hankey, J. M., Foley, J., & Ebe, K. (2015). Creation of the Naturalistic Engagement in Secondary Tasks (NEST) distracted driving dataset. *Journal of Safety Research*, 54, e29-36.
- Perez, M. A., Angell, L. S., & Hankey, J. M. (2015). Assessment of naturalistic use patterns of advanced infotainment systems. *Human Factors*, 7, 674-688.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Seppelt, B.D., Seaman, S., Lee, J., Angell, L.S., Mehler, B. & Reimer, B. (2016). *Glass half-full: Predicting crashes from near-crashes in the 100-Car data using on-road glance metrics*. Manuscript submitted for publication.
- Seaman, S., Lee, J., Angell, L., Mehler, B., Seppelt, B., & Reimer, B. (2016, October). Exploring Generalizability of Field Experiment Radio Tasks with Naturalistic Driving Data: A Comparison with SHRP2 NEST. In *Proc. of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct* (pp. 111-116).
- Victor, T., Dozza, M., Bårgman, J., Boda, C. N., Engström, J., & Markkula, G. (2014). *Analysis of Naturalistic Driving Study Data: Safer Glances, Driver Inattention, and Crash Risk*. (SHRP 2 Safety Project S08A). Gothenburg: Safer Vehicle and Traffic Safety Centre.
- Yekhshatyan, L. (2010). *Detecting Distraction and Degraded Driver Performance with Visual Behavior Metrics*. (Unpublished doctoral dissertation). University of Iowa, Iowa City, IA.