IN THE CONTEXT OF WHOLE TRIPS: NEW INSIGHTS INTO DRIVER MANAGEMENT OF ATTENTION AND TASKS

Linda Angell[1,], Sean Seaman[1], Rashmi Payyanadan[1], Wayne Biever[1], Bobbie Seppelt[2], Bruce Mehler[2], Bryan Reimer[2]

[1]Touchstone Evaluations, Inc., Detroit, MI, USA
[2]MIT AgeLab & N.E. Univ. Transportation Center, Cambridge, MA, USA
Email: langell@touchstoneevaluations.com

Summary: It is becoming increasingly important to understand how drivers strategically manage tasks and thread attention across time, as they drive through varying situations and conditions -- and as they have the opportunity to delegate tasks to vehicle automation while taking up other tasks themselves. To develop an understanding of these higher-level driver behaviors requires a research focus on longer periods of driving -- even on "whole trip" driving. It may also require new tools and methods. Therefore, to explore insights and implications of a "whole trip" focus, data from 10 drivers were analyzed using methods tailored for identifying patterns within larger sequences of driving data than single-task epochs. The results are reported, discussed, and contrasted with more conventional approaches based on single-task epochs.

INTRODUCTION

Within the changing transportation system, the role of the driver is being transformed. Driving is evolving from simple control to a dynamic mix of controls, planning, and supervision of automated processes. Executive attention functions for managing tasks and information during driving are increasingly salient. However, studying this threading of tasks and information by drivers across time -- across moments of driving -- requires a broader perspective and a time window that is longer than individual tasks. It requires a consideration of periods large enough that driver transitions between activities can be observed, along with shifts in attention between tasks. These periods of time may even need to encompass "whole trips."

Driving data have been collected and available on whole trips for some time (e.g., Dingus et al., 2006). And while recent theories of attention and task management (e.g., the Goal Activation Model, Altman & Trafton, 2002, and the Threaded Attention Model, Salvucci & Taatgen, 2008) suggest that there may be value in examining longer periods of driver interaction (to observe transitions between tasks or points of attention) -- it has been difficult to process and analyze long periods of driving data. Until now, available analysis tools and methods (e.g., glance coding methods) have been so resource-intensive that applying them to entire trips of data has often been precluded. This has led to targeted application of tools and methods to small "snapshots" of time, extracted from longer trips -- and to particularly salient periods of time for study (e.g., pre-crash periods).

However, as the rate of change within the transportation system increases, there is increasing urgency to look beyond small snapshots of time in understanding driver behavior. And this will

likely require a different vantage point on naturalistic data -- one that encompasses continuous data containing multiple periods of activities and tasks, uses a more wholistic analysis approach, and new tools that can examine longer data sequences in identifying patterns and time-course changes within them.

Therefore, the objective of this research was to explore the analysis of driver behavior across "whole trips" and to illustrate the types of findings that can emerge about ***engagement in secondary tasks during extended drives.*** This paper also contrasts findings from the "whole trip" approach with findings produced by more traditional analysis of small epochs or snapshots of driving that are defined by single tasks. For this comparison, driver interactions with the car stereo were selected as a focus.

METHOD

A dataset was analyzed using "whole-trip" analysis techniques to identify the insights and implications of these approaches -- and to compare the results with findings reported in published work from "single-task epoch-based" approaches.

Dataset. The driving behaviour data used here were acquired from 10 research participants who drove a Volvo S90 in the Boston-area for 4 weeks, using the instrumentation and methods described by Fridman et al. (2017). The full dataset contained approximately 2.5 million rows of time-synchronized data and included (among other data):
   • Location and duration of glances for each driver, manually-annotated from video
   • Secondary task activity for each driver, manually-annotated from video
   • Vehicle velocity for each driver (recorded by the data acquisition system)
The analyses conducted used only a subset of these data that pertained to the **driver's use of the car stereo** during "highway drives." All raw data were recorded at 30Hz. For this work, however, data were aggregated to 1s intervals after determining that the computational resources required for running analyses would become feasible at that sampling rate, while still preserving the key behaviours.

Vehicle Interfaces. One of two identical instrumented vehicles were used by all participants for their driving. The vehicle's interface provided a set of steering wheel (SW) buttons to the right of the steering wheel (SW) hub which could be used for interactions with the car stereo. When the controls were operated, a display in the lower right corner of the instrument cluster displayed information related to stereo-tasks. The center stack also allowed interaction with the stereo. Driver interactions with controls, as well as secondary task activities, and glances, were recorded on video (and manually coded from video). The task activities that were coded included: Holding Phone, Manipulating Phone, Handheld Conversation, Handsfree Conversation, Voice Navigation, Voice Phone, Steering Wheel (SW) Stereo, Steering Wheel (SW) Phone, Center Stack Climate, Center Stack Phone, Center Stack Navigation, Hands Off Wheel, and Other.

Analysis Techniques for Whole Trips

Generalized Additive Models (GAMs) were used to analyze the patterns of behavior within the highway drives. GAMs are a type of regression model which allow the effects of multiple

variables to be considered at once -- and are uniquely suited for estimating unbiased non-linear effects (Hastie & Tibshirani, 1987).

Normalization of whole trip times to a common scale. Each trip made by a driver lasted for a different length of time. Therefore, to compare across trips (for example, to compare the "beginnings," "middles," and "ends" of trips) -- and to enable the meaningful application of GAMs -- a normalization procedure was used to adjust all trips to a common scale (see Figure 1). Each time interval within a trip was divided by total trip time for that trip. This resulted in the bounding of trip times between 0 and 1 -- and placed them on a scale that could be interpreted as representing **percent of time into trip**.
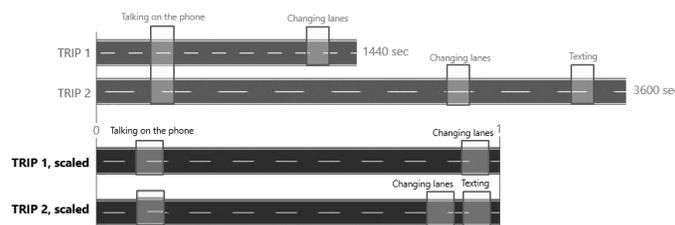

Figure 1. Illustration of scaling undertaken for highway trip times. Top panel (in lighter gray) shows original unscaled trip lengths; bottom panel shows trip lengths normalized to a common scale.

In addition, once GAM analyses were completed, **'ribbon charts' or 'timeline charts'** of secondary tasks and glances were created to visually depict the patterns of activity at the level of individual drivers. These plots were made using the data at the originally sampled rate of 30 Hz.

Analysis Techniques for Single Task Epochs. Results reported below for the single task approach were drawn from published and/or archived data in which task durations and glance metrics were computed following standard methods for individual task epochs.

RESULTS

Whole Trip Analyses
Trip-level patterns of task activity and glance behavior were revealed by the GAM analysis. These are shown in Figure 2, which shows *changes in task activity* on the car stereo (via the SW buttons), as related to *changes in glances to the instrument cluster*. These changes are represented by waveforms with peaks and troughs that recur throughout drivers' trips.
Effect of manipulating the steering-wheel stereo controls on glances to the instrument cluster. Figure 2 shows two patterns that emerged in the relationship between driver interactions with the steering-wheel-stereo-controls and concurrent glances to the instrument cluster (where task-related information was displayed). The first of these patterns is highlighted in the green dashed-line boxes of Figure 2 which identify recurring single peaks of increased activity on the steering-wheel stereo controls -- that were associated with a concurrent *significant increase in glances* to the instrument cluster at each corresponding point in the trip. (In other words, for these green dashed boxes, when there was at least a 1% increase in activity with the stereo via the steering-wheel buttons, there was also a significant increase in glances to the cluster area <u>during</u> those button presses). A second pattern emerged as well. It is highlighted in the purple solid-line boxes. This pattern consisted of a recurring couplet of single and double peaks for stereo activity (showing increased activity on the steering-wheel buttons) -- with a corresponding *decrease or*

total absence of glances to the instrument cluster area. The "purple solid-line box" activity would thus appear to be a different task activity than the one highlighted in the dashed boxes -- one that can be accomplished "blindly," without looking at the controls or instrument cluster (and/or one in which the driver's finger or thumb was already in position, based on prior glances), allowing the driver to look elsewhere when manipulating the steering-wheel buttons.
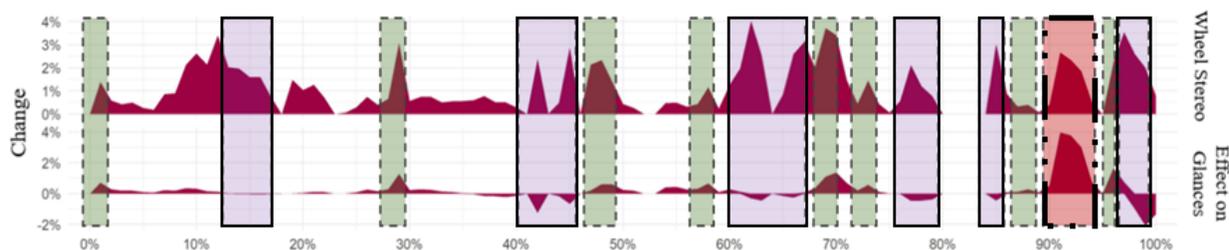

Figure 2. GAM analysis results showing Percent Change (in two variables) as a function of Percent of Time into Trip (along the horizontal axis). The relationship between changes in the two variables is shown for each moment in time across highway driving trips, with % change in Activity on SW Stereo Buttons <u>on top</u>, and % change in Glances to the Instrument Cluster <u>on bottom</u>, both at a sampling rate of 1 Hz. In the top panel, the *y*-axis represents the change from a *baseline* of zero (no SW stereo task activity). In the bottom panel, the *y*-axis represents the deviation from *baseline* glance behavior associated with a specific task activity. For this, the *baseline* glance behavior is determined by GAM analyses based on percentage of time into the trip, and the absence of any task activity (i.e. all task activities including SW stereo task activity). This baseline can be non-zero.

One outlier in the pattern of the glances to the instrument cluster can be seen at 90% into the trip (red dash-dot box) glances to the instrument cluster sharply increased during interaction with the steering-wheel stereo buttons. Since this is near the end-of-the-trip, it may be associated with the period when drivers are preparing to exit the highway -- and may reflect co-occurring glances to information in the cluster beyond the stereo task (e.g., to the speedometer or other information).

Ribbon plots (Figure 3) provide a "close-up" view for the actual task activity of one participant out of the cohort shown in the GAM results. Such plots can be useful for confirming as well as gaining further insight into patterns revealed by the GAM analyses. The first plot (**Plot 3A**) depicts the time history (on the x-axis) of all activities and glance patterns for this participant's entire highway driving trip (~1 hour). Each row ("ribbon") on the y-axis shows engagement in a different secondary task activity or glance location. This driver carried out tasks with the climate controls, smartphone, and stereo (using both center stack and SW controls) -- and, in doing these tasks, interleaved glances between the road, the instrument cluster, and other locations within the vehicle. The red box frames the two rows which depict: (1) the use of the SW stereo buttons and (2) glances made to the instrument cluster. The three other ribbon plots in Figures 3B.1, 3B.2, and 3B.3 each provide an *enlargement* of one of the 7 stereo interactions shown in Plot 3A that occurred during the sample participant's trip. They illustrate the different types of attentional patterns revealed by the GAM analysis -- and are provided in the order in which they occurred in the participant's time history for the trip. **Plot 3B.3** shows a task that is 15s long -- which occurred at the end of the highway trip -- and required 1 leading glance prior to the task, and 4 glances during the task -- all of which alternated with healthy-sized glances to the road. This task exemplifies the type of task identified by the first pattern in the GAM analysis. **Plot 3B.2** shows a task that is about 12s long -- which occurred 90% of the way into the highway trip (prior to exiting) -- and required 6 glances to be made during the task, as well as 2 glances immediately following (perhaps to verify settings). These glances varied from 1s to 2s in length, with total

glance time to complete the stereo task summing to about 7s total, and an additional ~3s afterward. The plot also reveals that a **smartphone task preceded this stereo task** (purple ribbon at top), **and partially overlapped with it** -- such that glances to the phone (shown on "internal glances" ribbon) and road were interleaved with the stereo task, and the lengths of glances to the road were shortened (bottom-most ribbon). This fell within the window identified by GAM as showing unusually high glance activity prior to highway exit. **Plot 3B.1,** in contrast, shows a task of a different type (the one identified by the second pattern in the GAM analysis). This task occurred earlier in the trip -- was very brief -- about 2s in length (likely a volume adjustment or other single button press task) -- wherein no glance was made to the control or cluster area during the button press -- but a single short glance (~200 ms in length) was made to the cluster area following the interaction. Thus, these panels each show different types of task interactions performed with the stereo at different times while driving.
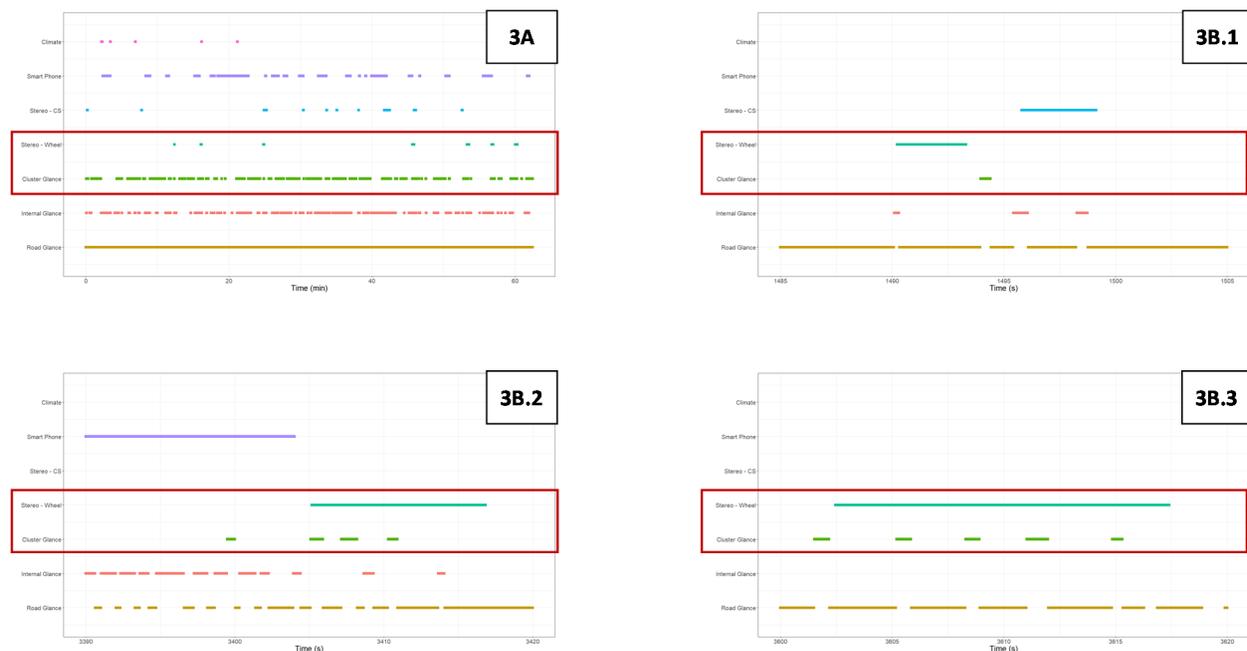

Figure 3. Ribbon plots of secondary task activities and glances over time for one driver, illustrated in four panels. Panel 3A shows the <u>whole trip</u> in its entirety. (Note: On the ribbon for road glances, individual glances to the road were so numerous that they become indiscernible when the plot is reduced in size to the degree shown here, and they are squeezed together, making gaze *appear to be* continuous for this location -- though it was not, as is apparent in the enlarged figures of 3B.1, 2, 3). The other panels show three different individual interactions with the stereo selected from 3A, each zoomed in to enlarge one stereo interaction. 3B.1 shows a short stereo interaction that is followed by a short glance. 3B.2 shows a longer stereo task requiring multiple glances to accomplish that follows a smartphone task. 3B.3 shows another long stereo task, done with several glances. The horizontal axis of all plots displays the timeline for each (it is about 10-20s in length for B.1-B.3). The vertical axis of each plot provides a "row" for each "ribbon" of activity, as follows (in order from top): Climate, Smart Phone, Stereo-Center Stack, Stereo-SW, Cluster-Glance, Internal-Glance, and Road-Glance. Red boxes highlight the rows showing stereo button activity & glances to the instrument cluster.

Single Epoch-Based Analyses

As a comparison with the findings from whole-trip analyses, consider evaluations of secondary task effects on drivers in which individual tasks are tested (e.g., under distraction guidelines such as NHTSA's or the Alliance's). Such evaluations yield performance metrics for each task tested

(see Figure 4, drawn from Reimer & Mehler, 2013-18A). Data of this type can be very useful in determining how demanding a task is of a driver's resources -- and whether its interface could benefit from improvement prior to release in a product -- which is its purpose. The guidelines have established common methods that can be used throughout the industry for task testing. *However,* this single-task approach is not designed to be informative about the various strategies that drivers may use for performing a task -- nor to provide insight about the conditions under which the task may be initiated, whether the task will ever be performed by actual drivers on the road, nor whether it will be performed in combination with other tasks. Actual use patterns are not reflected in the testing approach, nor are they intended to be at this time. Thus, much remains to be learned about how drivers manage transitions from task-to-task, and thread attention between tasks and the road. It is in this context that analysis of natural whole trip data is needed.

Figure 4. Reprinted from Reimer and Mehler (2013-18A). Total off-road glance time for each task, with the 12 second threshold from the NHTSA Phase 1 Guidelines shown as a dashed line. The individual line segments above each bar represent the 85th percentile point in the sample distribution for each task. One outlier data point in the Nav Entry task is excluded from view to improve the readability of the plot. Note that the NHTSA threshold values are shown here for discussion purposes only -- since, among other considerations, the sample does not conform to the NHTSA recommended age distribution and the data were collected under real driving conditions as opposed to the specified simulation conditions. Arrows (superimposed on the figure) highlight radio tuning tasks: radio easy manual, radio easy voice, radio hard voice, and radio hard manual tasks (from l to r).
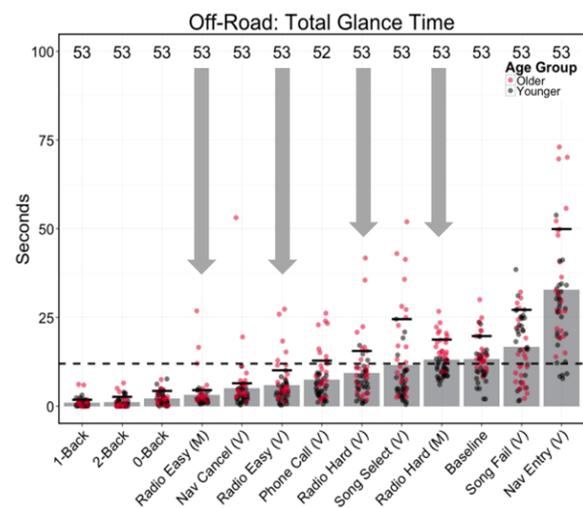
DISCUSSION

More efficient manual annotation tools and AI-fueled approaches are beginning to facilitate deeper investigations of how driving and non-driving activities are threaded across trips taken over weeks and months of driving and allowing researchers to begin looking beyond the demand of single task interactions isolated from the broader strategic aspects of real-world engagement. In addition, new analytic tools (like GAM) offer the capability for discovering patterns in data across longer temporal sequences. The GAM analysis here provided a trip-level picture of stereo-system use via SW buttons during highway drives for a sample of 10 drivers. Additionally, the ribbon plots provided a view of the interactions carried out across a trip by one specific participant within the sample. Together, these analyses provided one illustration of the types of insights that can emerge from a consideration of driver task management using a more "telescopic" view of a whole trip than a "microscopic" view of single-task epochs.

Interestingly, patterns obtained from the whole-trip approach are consistent with findings from a small number of studies in the literature on naturalistic use of infotainment systems (e.g., Perez, Angell, & Hankey, 2014) -- while still providing new insights. For example, Perez et al. (2014) observed that drivers spent 99% of their driving time listening to the infotainment system and adjusted the infotainment system at the average rate of 6.1 interactions per hour (vs. 7 times/hour for the participant illustrated here). Further, mean eyes-off-road time fell below 10 s for all

interactions. The ribbon charts for the participant here show total glance times for the longest interaction to be in the range of 7 - 10s), which corresponds well with Perez et al. (2014), as well as with radio task data in Reimer & Mehler (2013). Thus, the consistency of the new results with the prior ones gives initial confirmation of the validity of patterns from the new analytic tools.

Additionally, new and unique insights emerged from the whole trip analysis approach. These included: (1) T*he identification of transitions within trips* as key points upon which to focus more research attention; (2) *The fact that relationships between certain tasks may be important,* in terms of how they are handled by drivers -- e.g., coordinating music selection from both a smartphone app and the car stereo while driving; and (3) *The fact that attention and task shifting are frequent across trips -- and are characterized by discernible patterns*. With a trip-level focus, advanced analytics (such as GAM and other tools) offer the possibility for extracting such patterns from longer sequences of driving data. This can, in turn, advance the understanding of how drivers manage tasks over time as a function of driving situations, risk, and automation.

ACKNOWLEDGMENTS

REFERENCES

Altmann, E. M. and Trafton, J. G. (2002). Memory for goals: An activation-based model. *Cognitive science*, 26 (1), 39-83.

Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., Perez, M. A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z. R., Jermeland, J., and Knipling, R.R. (2006). *The 100-Car Naturalistic Driving Study: Phase II – Results of the 100-Car Field Experiment*. Report No. DOT HS 810 593. Wash., D.C.: NHTSA.

Fridman, L., Brown, D., Glazer, M., Angell, W., Dodd, S., Jenik, B., Terwilliger, J., Kindelsberger, J., Ding, L., Seaman, S., Abraham, H., Mehler, A., Sipperley, A., Pettinato, A., Seppelt, B., Angell, L., Mehler, B., & Reimer, B. (2017). *MIT Autonomous Vehicle Technology Study: Large-Scale Deep Learning Based Analysis of Driver Behavior and Interaction with Automation*. Massachusetts Institute of Technology, Cambridge, MA. https://arxiv.org/pdf/1711.06976.pdf

Hastie, T., & Tibshirani, R. (1987). Generalized additive models: some applications. Journal of the American Statistical Association, 82(398), 371-386.

Perez, M. A., Angell, L. S., & Hankey, J. M. Assessment of naturalistic use patterns of advanced infotainment systems. *Human Factors*, 57, 4, p. 674-688.

Reimer, B. & Mehler, B. (2013). *The effects of a production level "voice-command" interface on driver behavior: summary of findings on reported workload, physiology, visual attention, and driving performance.* MIT AgeLab White Paper No. 2013-18A. Massachusetts Institute of Technology, Cambridge, MA.

Salvucci. D. D. and Taatgen, N.A. (2008). Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review*, 115 (1), 101.