ON-ROAD EVALUATION OF DESTINATION ENTRY AND WAYFINDING TASKS: COMPARISONS AGAINST NORMAL DRIVING

Miguel Perez[1], Derek Viita[1], Jonathan Hankey[1], Sherri Voran-Nowak[2], & Steven Tengler[2]
[1] Virginia Tech Transportation Institute (VTTI), Blacksburg, Virginia, USA
[2] OnStar Corporation, Detroit, Michigan, USA
Email: mperez@vt.edu

Summary: While relative comparisons between "distracting" tasks (e.g. dialing a cell phone vs. talking on the cell phone) are useful, "normal driving" remains the benchmark for any task performed by the driver while a vehicle is in motion. Arguably, tasks that are less risky will result in observed patterns of driver behavior that are closer to those observed during normal driving. This paper describes the outcome of a study to compare destination entry and wayfinding across different navigation devices (with different input modalities) against epochs where the driver was not tasked with any other secondary or tertiary tasks (beyond occasional conversation with the experimenter). Results indicate some significant differences between destination entry tasks and normal driving, the magnitudes of which are mainly modulated by the input modality. Differences were less obvious during the navigation tasks, likely due to the intermittent nature of interactions with the navigation device in that context. Total eyes off-road time was also subjected to comparisons against previously published crash and near-crash risk estimate models. The results suggest that, assuming confidence in the models, there may be differences in the levels of crash and near-crash risk associated with different navigation devices. The approach is presented as a potential additional metric to consider in assessing devices that are used by drivers in moving vehicles.

INTRODUCTION

Recent advances in technology have allowed the inclusion of navigation systems in a variety of platforms, ranging from embedded in-vehicle devices all the way to applications designed to run on nomadic devices. These same advances have also allowed the use of multiple modalities to perform destination entry. As these systems become more prevalent, especially those in devices not specifically designed for in-vehicle use, it is important to understand the effects that using these systems may have on drivers that choose to do so while operating a moving vehicle.

A recently published study (Perez, Viita, Hankey, Voran-Nowak, & Tengler, 2011) documented driver behaviors associated with destination entry and wayfinding tasks on different navigation systems. While there were no statistical differences in the driving performance measures evaluated across the different navigation systems, there were significant differences in the eye glance patterns that were observed. Voice-controlled destination entry tasks required less glances away from the forward roadway and driving-related locations than tasks using visual-manual controls. During navigation, visual displays with more information density were associated with higher visual demand. Furthermore, there were advantages identified in navigating with any of the devices tested in comparison to navigating with the traditional method of printed directions.

In spite of the useful information they provide, relative comparisons such as Perez, et al. (2011) are limited in that there is no assessment of the level of crash or near-crash risk associated with any particular task. Crash or near-crash risk assessment has historically been limited by the data available, and the estimates themselves are often the source of substantial debate. While epidemiological methods for risk assessment are well developed and established they are often difficult to apply in the transportation field due to the lack of necessary data, at least until the recent advent of large-scale naturalistic driving studies. This does not imply that there have not been attempts to apply these methods (e.g., Mcevoy, et al., 2005; Redelmeier & Tibshirani, 1997), simply that the estimates provided by these studies are often controversial since small biases in the data selection can produce sizeable variation and inaccuracy in the calculated risk.

Results from large scale naturalistic studies have been used to produce estimates of crash and near-crash risk. These datasets record actual driving and are not limited by the availability of records, witnesses, and other factors that can bias other data sources. Two notable examples of crash and near-crash risk estimation based on naturalistic data (Klauer, Dingus, Neale, Sudweeks, & Ramsey, 2006; Klauer, Guo, Sudweeks, & Dingus, 2010) are based on data from the 100-Car study (Dingus, et al., 2006). These approaches applied epidemiological methods to carefully selected samples from naturalistic driving data to produce estimates of the likelihood that particular driver behaviors could result in a crash or near-crash, using sample epochs of "normal driving" as the baseline in the assessment. Klauer, et al. (2010), particularly, has produced the most reliable links between eye glance behavior, especially the total eyes off-road time (TEORT) measure, and crash and near-crash risk to date, at least until larger-scale naturalistic datasets (e.g., the Naturalistic Driving Study from the Strategic Highway Research Partners II program) become available. They examined eyes off-road time over a period of 15 sec. for a variety of tasks that resulted in a crash or near-crash, and contrasted them against "baseline" samples selected using a case-crossover methodology. The result is a model that estimates the risk of a crash or near-crash based on how long the eyes are diverted from the forward roadway over a period of 15 seconds. In turn, the availability of this information may allow for the extrapolation of this risk to other tasks (and their visual demands) which was not part of the original sample.

This investigation was completed to compare the destination entry and navigation tasks reported in Perez, et al. (2011) to a "normal driving" baseline. In addition, the results were overlaid on the crash and near-crash risk estimate model generated by Klauer, et al. (2010) with the goal of illustrating a novel approach to estimate risk from secondary and tertiary tasks.

METHOD

Details about the experimental method are described in Perez, et al. (2011), and summarized here. Twenty-four participants (25 to 55 years old), equally distributed by gender, drove an instrumented 2010 Buick LaCrosse on public roads. The experiment was conducted on five different navigation systems/approaches: VZ Navigator™ (cell phone-based application, used in visual-manual mode), Tom Tom™ GO 740 LIVE (portable navigation device, used in visual-manual mode), OnStar™ Turn-By-Turn (embedded system, voice-controlled), OnStar™ Destination Download (embedded system, mostly voice-controlled with some visual-manual interaction), and traditional printed directions. Experimental trials consisted of the participant

being trained statically on destination entry for a device (with the exception of printed directions), driving to a low-density rural highway, completing the destination entry task while driving on that highway, and following the directions to the destination provided by the device.

Destination entry and navigation portions of the task for each device were isolated and used in the original analyses. The current investigation isolated two additional segments of time for each participant drive. The first segment consisted of driving data gathered between the time when the last navigation task was complete and the participant's return to their point of origin. This driving period occurred in the same highway used for destination entry tasks, and is therefore directly comparable to those segments. During this period, the participant had no other task than to drive back to their point of origin. Since this period was defined *post hoc*, casual conversation between the experimenter and the participant was not limited and occurred in most cases. The second segment consisted of driving data gathered at one randomly-selected interval between a destination being reached and the return to the highway to begin the next destination entry task. This driving period occurred in roads similar to those used for the navigation portion of the study, and therefore was used as a "normal driving" comparison for those segments. Casual conversation between experimenter and participant was not limited and occurred often.

The duration of the destination entry baselines was set at 120 seconds, as this was close to the average length of destination entry tasks. The duration of the navigation baselines depended on the segment that was randomly selected.

Data Analysis

All of the dependent measures described in Perez, et al. (2011) were calculated for both baseline segments, and driver performance with the various systems contrasted to these "normal driving" epochs through analysis of variance and *post hoc* comparisons (Tukey adjustment). A Type I error limit of 0.05 was selected.

Each of the destination entry and navigation tasks, along with their associated baselines, was further analyzed in terms of their TEORT. The goal of the analysis was to find a measure that was comparable to the model of crash and near-crash risk proposed by Klauer, et al. (2010). This model assumes maximum task duration of 15 seconds. To make the current results comparable, it was necessary to select a sample of experimental tasks and baselines (whose total duration was well in excess of 15 seconds) that was representative and meaningful when compared to Klauer, et al.'s model. The sample discussed in this paper was selected amongst several options (e.g. averaging TEORT over incremental 15-sec. segments, or over randomly selected samples that lasted 15 sec.). In the sample selected for analysis, TEORT was calculated over all possible 15-sec. segments (limited by the 10-Hz data collection rate) available on each task completed by each participant. From this set of TEORT figures, two were selected to represent each task performed by each participant in the analysis: the maximum and minimum TEORT. These two numbers were considered to represent the worst- and best-case scenarios, in terms of TEORT, for each task. In that sense, they can be conceptualized to represent risk "limits." To the extent that significant differences in risk estimates are existent for either of these figures, differences in risk between tasks can be inferred. The Friedman test was used to detect statistically significant differences in maximum and minimum TEORT.

RESULTS

None of the kinematic measures showed statistically significant differences between the destination entry or navigation baselines and the driving segments where destination entry and navigation tasks were completed. In terms of duration, the baselines for navigation tended to be shorter than the duration of the tasks themselves (Baseline driving: 296.1 sec. (SD=108.1), Navigation tasks: 460.2 sec. (SD=170.1)). This difference was statistically significant for the OnStar™ systems, which showed longer task durations (although they were not statistically different from the other systems).

Eye glance analysis yielded a sizeable number of statistically significant differences between destination entry tasks and their associated baseline (Table 1). In general, the voice-controlled systems were similar or slightly different from baseline driving, whereas the visual-manual control systems tended to be very different from baseline driving. While the percentage of forward glances tended to remain the same across systems, the duration of these glances was reduced, especially for the Tom Tom™ and VZ Navigator™ systems. The frequency and duration of non-forward driving-related glances (e.g. glances to rear-view mirrors) were also reduced for these two systems. Non-forward non-driving-related glances (e.g., to the navigation system) occurred more often and were longer during destination entry tasks than for the associated baseline driving period.

Differences in eye glance behavior were less evident between navigation tasks and their associated baselines (Table 2). Generally, the OnStar™ Turn by Turn and VZ Navigator™ systems were statistically similar to the baseline condition, with the remaining systems being significantly different (statistical similarity was observed for some measures, however). "Practical" differences (e.g., beyond a few percentage points, or several seconds) for navigation were also less evident than during destination entry.

The analysis of the 15-second segments with maximum and minimum TEORT showed statistically significant differences in those measures as a function of System during destination entry and navigation segments. TEORT values were also superimposed in the Klauer et al. (2010) crash model plots (Figure 1). The Friedman test showed a significant effect of System for maximum TEORT during destination entry ($p<0.0001$), minimum TEORT during destination entry ($p<0.0001$), and minimum TEORT for during navigation ($p=0.0033$). This last effect, however, was deemed to be of little practical significance (Figure 1d) and not analyzed further.

Post hoc tests for maximum TEORT in destination entry tasks showed significant differences between all systems and baseline driving, and between all systems except for the Tom Tom™ and VZ Navigator™. Minimum TEORT for destination entry differed significantly between the baseline condition and all systems. For this measure, however, OnStar™ Destination Download and OnStar™ Turn by Turn had statistically similar minimum TEORT values, as did the Tom Tom™ and VZ Navigator™ systems. The minimum TEORTs for these two groups of systems, however, were significantly different from each other (Figure 1c).

DISCUSSION

As found in Perez, et al. (2011) the vehicle performance measures did not indicate any difference between the destination entry and navigation tasks and their associated baselines. Observation of average values also failed to suggest any trends in the data that may become significant with increased statistical power. This in turn suggests that the vehicle performance measures used in this study were relatively insensitive to the varying levels of driver workload that participants were subjected to, a finding that agrees with some previous literature on distraction (see Collet, Guillot, & Petit, 2010 for a review)

Task duration for baseline navigation tasks, which was not pre-determined, was significantly shorter than task durations for the actual navigation tasks. This was very likely an artifact of the routes selected. The experimenter guided the participant back to the low-density highway for the next task, and the route chosen was one that minimized the time to begin the next task, rather than one that took the participant back to the starting point. The types of roads traversed and maneuvers completed were similar, however, in spite of the difference in routes and durations.

Table 1. Average eye glance results across participants during destination entry and *post hoc* test results

		Average Values					Post hoc Test Results			
		Baseline (Reference Value)	Destination Download	Tom Tom	Turn By Turn	VZ Navigator	Destination Download	Tom Tom	Turn By Turn	VZ Navigator
Percent of Glances (based on frequency)	Forward	47.0	49.2	47.7	48.3	47.8	>			
	Not Forward - Driving-Related	37.1	17.5	4.2	21.1	5.0	<	<<	<	<<
	Not Forward - Not Driving-Related	15.9	33.4	48.1	30.6	47.2	>	>>	>	>>
	Navigation System Only		*22.2*	*43.1*	*14.7*	*43.0*				
Percent of Glances (based on duration)	Forward	84.5	70.7	35.1	76.6	42.8	<<	<<<<	<	<<<
	Not Forward - Driving-Related	10.8	9.3	2.5	10.1	3.7		<		<
	Not Forward - Not Driving-Related	4.6	19.9	62.3	13.2	53.5	>>	>>>>	>	>>>
	Navigation System Only		*13.9*	*59.0*	*6.5*	*50.8*				
Glance Rate (glances/min)	Forward	12.3	20.7	30.8	18.0	31.2	<	<<	<	<<
	Not Forward - Driving-Related	9.5	7.6	2.7	8.3	3.3		<		<
	Not Forward - Not Driving-Related	3.9	13.9	31.1	11.2	30.7	>	>>	>	>>
	Navigation System Only		*9.4*	*27.8*	*5.8*	*28.0*				
Total Glance Duration (sec)	Forward	83.7	82.0	35.3	73.4	62.1		<		<
	Not Forward - Driving-Related	11.4	10.6	2.6	9.9	5.3		<		<
	Not Forward - Not Driving-Related	4.6	23.2	65.5	13.0	80.2		>		>
	Navigation System Only		*16.0*	*61.8*	*6.4*	*77.0*				
Average Glance Duration (sec)	Forward	5.9	2.2	0.7	3.1	0.8	<	<<	<	<<
	Not Forward - Driving-Related	0.7	0.7	0.6	0.8	0.7				
	Not Forward - Not Driving-Related	0.7	0.9	1.2	0.7	1.1	>	>>>		>>
	Navigation System Only		*0.9*	*1.3*	*0.7*	*1.1*				
** Percent of Glances over 2 sec - Not Driving-Related		0.2	0.8	5.3	0.0	2.2		>		
** Longest Glance Duration - Not Driving-Related (sec)		1.0	2.0	2.9	1.4	2.8	>	>		>

>, < indicate statistically significant difference between the system and the baseline, symbols are repeated to indicate the strength of the difference (e.g., << indicates that the value is smaller than the corresponding value for the baseline and smaller than the value for the system with the < sign in the row); Blank space - No statistically significant difference observed between the system and the baseline.

** - Note that the numbers in these rows (and the rest of the table) are averages across participants. In the baseline condition, there is some percentage of "Not Driving-Related" glances although the value for "Longest Glance Duration - Not Driving-Related) is less than 2 sec. This is because there were a few participants with longest not driving-related glances greater than 2 sec in this condition (making the **average** percentage greater than zero), but these longest glances were not sufficient to result in an **average** longest not-driving related glance duration greater than 2 sec.

Table 2. Average eye glance results across participants during navigation and *post hoc* test results

		Average Values						Post hoc Test Results				
		Baseline (Reference Value)	Destination Download	Printed Directions	Tom Tom	Turn By Turn	VZ Navigator	Destination Download	Printed Directions	Tom Tom	Turn By Turn	VZ Navigator
Percent of Glances (based on freq.)	Forward	49.7	49.3	49.4	48.9	49.3	49.6				<	
	Not Forward	50.1	50.7	50.6	51.0	50.6	50.4	>	>			
	Navigation System Only		*23.1*	*28.5*	*28.3*	*6.1*	*12.8*					
Percent of Glances (based on duration)	Forward	83.1	76.4	71.2	75.7	83.4	80.4	<	<	<		
	Not Forward	16.8	23.6	28.8	24.5	16.6	20.0	>	>	>		
	Navigation System Only		*11.9*	*17.4*	*13.8*	*1.8*	*5.5*					
Glance Rate (glances/min)	Forward	9.2	13.0	12.7	13.0	9.1	10.7	>	>	>		
	Not Forward	9.2	13.4	13.0	13.4	9.4	10.8	>	>	>		
	Navigation System Only		*6.2*	*7.4*	*7.5*	*1.2*	*3.0*					
Average Glance Duration (sec)	Forward	6.5	3.9	3.6	3.8	6.5	5.1	<	<	<		
	Not Forward	1.2	1.1	1.3	1.1	1.1	1.1					
	Navigation System Only		*1.1*	*1.4*	*1.1*	*0.9*	*1.1*					
Percent of Glances over 2 sec - Not Driving-Related		4.4	2.7	5.7	3.9	3.8	2.8					
Longest Glance Duration - Not Driving-Related (sec)		7.5	5.4	8.4	5.4	6.1	8.5					

>, < indicate statistically significant difference between the system and the baseline, symbols are repeated to indicate the strength of the difference (e.g., << indicates that the value is smaller than the corresponding value for the baseline and smaller than the value for the system with the < sign in the row); Blank space - No statistically significant difference observed between the system and the baseline.

Eye glance analyses for destination entry and its baselines suggest that there were measurable differences between periods of driving concurrently with a destination entry task and periods of "normal" driving. These differences became more salient as the requirements from visual-manual task performance increased. While still different from "baseline" driving in some measures, the systems that allowed for voice entry of destination information resulted in less eye glance effects than systems that required visual-manual entry. Voice entry systems generally allowed for more and longer driving-related glances than visual-manual entry systems. Driving-related glances have been associated with decreased crash risk (Klauer, et al., 2006).

Eye glance analyses for navigation, in contrast, show smaller differences between system use and baseline driving. The eye glance analysis for navigation, however, did not distinguish between driving-related and non-driving-related non-forward glances. Therefore, this finding should be interpreted cautiously. The results suggest that, drivers in systems with more visual information accessed displays at the expense of briefer glances to the forward roadway. Driving-related glances may also be impacted, but the granularity of the glance locations does not allow such a conclusion to be reached with the current data.

Results for the TEORT analyses for destination entry indicated differences between the systems and baseline driving. The larger the amount of visual-manual interaction required to complete the task, the larger the difference between the system and baseline driving. Translated to estimated risk based on Klauer, et al. (2010), there seems to be a non-statistically significant trend for visual-manual control to place the driver at increased risk of a crash compared to voice control. No differences were observed in the navigation tasks, where interaction with the systems was sporadic and did not involve any manual entry. The true existence of these differences, and their magnitude, will require further data that tightens the confidence intervals for the risk estimates and considers elements of cognitive distraction in their assessment. Nonetheless, Klauer, et al.'s model, and the analysis approach presented here, appear to be useful in providing initial estimates of risk for these types of tasks and illustrating the differences inherent in different systems and control approaches.

(a)

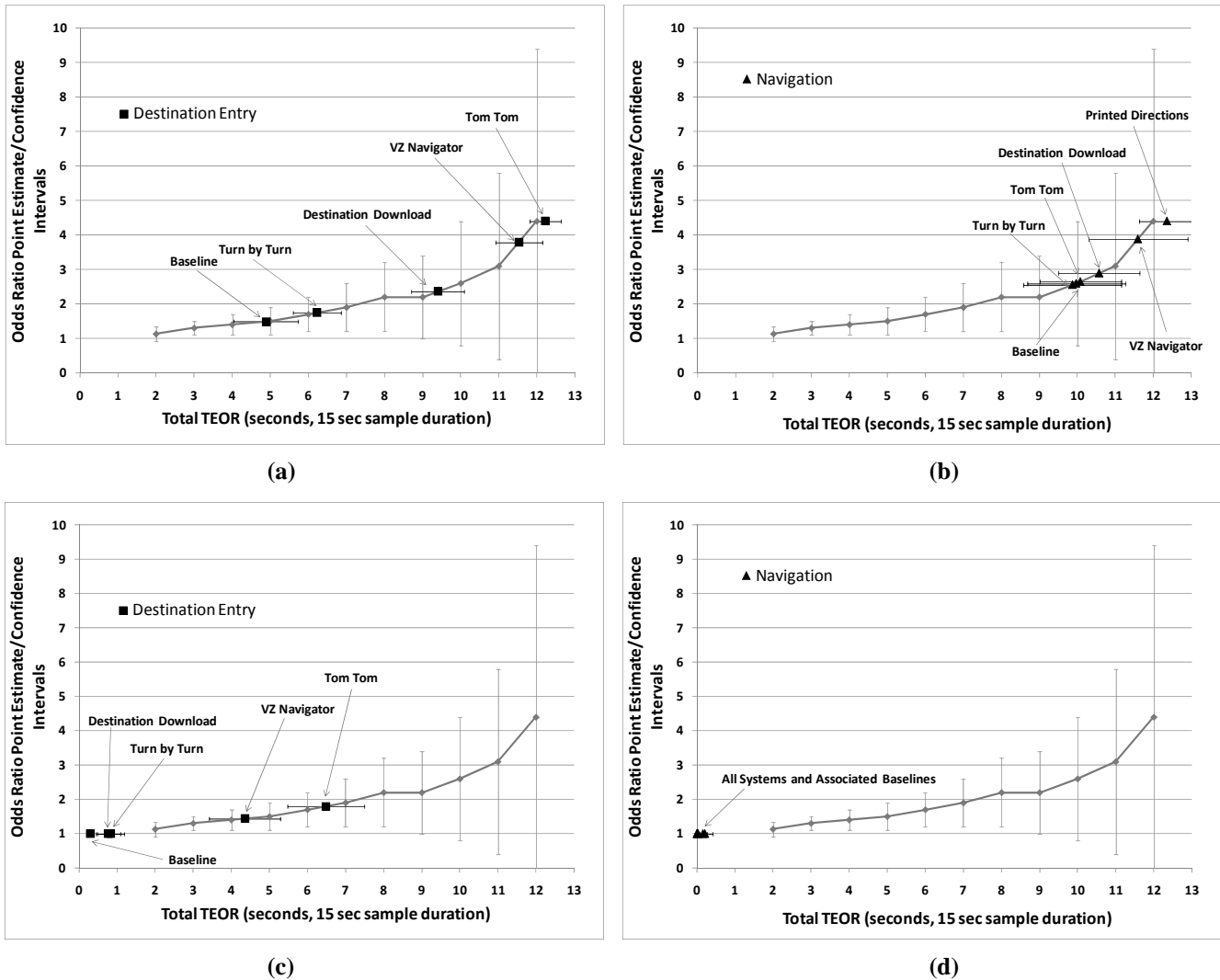

(b)

(c)

(d)

Figure 1. Average of maximum and minimum TEORT values over 15-second intervals for destination entry, navigation, and associated baselines; (a) Maximum for destination entry, (b) Maximum for navigation, (c) Minimum for destination entry, (d) Minimum for navigation
(values are overlaid on the Klauer et al. (2010) odds ratio model plots)

REFERENCES

Collet, C., Guillot, A., & Petit, C. (2010). Phoning while driving i: A review of epidemiological, psychological, behavioural and physiological studies. Ergonomics, 53(5), 589-601.

Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., et al. (2006). The 100 car naturalistic driving study, phase ii - results of the 100 car field experiment (Report No. DOT-HS-810-593). Washington, D.C.: National Highway Traffic Safety Administration.

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data (Report No. DOT HS 810 594). Washington, D.C.: National Highway Traffic Safety Administration.

Klauer, S. G., Guo, F., Sudweeks, J., & Dingus, T. A. (2010). An analysis of driver inattention using a case-crossover approach on 100-car data: Final report (Report No. DOT HS 811 334). Washington, D.C.: National Highway Traffic Safety Administration.

McEvoy, S. P., Stevenson, M. R., McCartt, A. T., Woodward, M., Haworth, C., Palamara, P., et al. (2005). Role of mobile phones in motor vehicle crashes resulting in hospital attendance: A case-crossover study. British Medical Journal.

Perez, M. A., Viita, D., Hankey, J. M., Voran-Nowak, S., & Tengler, S. C. (2011). Nomadic and embedded navigation system evaluation based on user performance in destination entry and navigation tasks. Paper presented at the Proceedings of the Transportation Research Board 90th Annual Meeting.

Redelmeier, D. A., & Tibshirani, R. J. (1997). Association between cellular-telephone calls and motor vehicle collisions. The New England Journal of Medicine, 336(7), 453-458.