A Driving Simulation Study Examining Destination Entry with iOS 5 Google Maps and a Garmin Portable GPS System

Celena Dopart[1], Anders Häggman[2], Cameron Thornberry[1],
Bruce Mehler[3], Jonathan Dobres[3], Bryan Reimer[3]

[1]Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA
[2]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA
[3]MIT AgeLab & New England University Transportation Center, Cambridge, MA

A simulation study compared 23 young adult drivers' task completion time, mean glance time, number of glances, and percentage of long glances while performing a navigation entry task with a Garmin portable GPS system and a mobile navigation application (iOS 5 Google Maps) on an iPod Touch. We compared participants' performance on these tasks using the National Highway Traffic Safety Administration (NHTSA) eye-glance acceptance criteria. We found that, irrespective of the device used, no participant was able to complete the task within the recommended total time window of 12 seconds. When entering a destination into the iOS interface, only 73.9% of the drivers meet the NHTSA criteria for long duration glances. With the Garmin system, 91.3% of the participants meet this criterion. All participants were able to maintain adequate mean off road glance durations. Finally, we compared the NHTSA recommended method of assessing all off road glances to more traditional methods of assessing glances only to the task interface. Differences between the two methods are discussed.

INTRODUCTION

Dashboard- or windshield-mounted portable GPS units in automobiles are increasingly being replaced by smartphones running applications with comparable functions. According to a Mobile Consumer Report from Oracle conducted in 2010, more than 50 percent of 3,000 worldwide survey respondents thought their smartphone would replace other electronic devices, including GPS devices, within five years. By the end of 2011, 24 percent of respondents had already done so (Oracle, 2011).

In recent years, it has become evident that America's growing technology engagement while behind the wheel carries the potential for serious safety risks. A number of studies have assessed the impact of smartphone applications on driving (Basacik, Reed, & Robbins, 2011), including those with touchscreen interfaces (Reimer et al., 2012; Samuel, Pollatsek, & Fisher, 2011) and the use of navigation smartphone applications (Lee & Cheng, 2010; Quaresma, 2012). Numerous studies have also assessed the effects of independent GPS devices on driving performance (Tijerina, Parmer, & Goodman, 1998; Jensen, Skov, & Thiruravichand, 2010).

The issue of smartphone and GPS use while driving is of particular concern for the National Highway Traffic Safety Administration (NHTSA) as part of the NHTSA phase II distraction guidelines pertaining to portable electronic device use while driving (NHTSA 2010). NHTSA's guidelines pertaining to visual-manual distraction for in-vehicle electronics specify a new set of criteria for eye glance evaluations conducted in driving simulators. These criteria are based entirely on eye glances away from the forward road scene (NHTSA 2012), in contrast to more traditional distraction evaluations proposed by the Alliance of Automobile Manufacturers (the Alliance) (Alliance, 2006) that evaluate glances only to task-related areas (controls and displays).

While studies have investigated the effects of GPS or smartphone data entry while driving, research directly comparing more modern devices is quite limited. This report aims to investigate differences in task completion time and driver glance behavior during destination entry tasks into both a Garmin GPS unit and iPod Touch iOS 5 "Google Maps" application. Comparisons between the devices are computed using the NHTSA eye-glance evaluation metrics as well as a more traditional "glance to device" approach.

METHODS

Participants

The sample population targeted was relatively young (20-34 years) to increase the likelihood of drawing subjects with high levels of experience using cell phone applications and GPS units. Participants were screened for eligibility, which required active drivers with 3+ years of holding a valid driver's license. Compensation of $40 was provided.

A total of 25 subjects were recruited. Two subjects were excluded from analysis for having task completion times and/or glance frequencies more than three standard deviations greater than the group mean, leaving 23 subjects (11 female). Average age of the sample was 25.9 years (SD=3.86). Age did not differ significantly by gender ($W = 968$, $p = .491$).

Apparatus

The driving simulator was a stationary 2001 Volkswagen Beetle situated in front of a projection screen with a 40 degree view of a virtual highway. The simulation's graphical updates were generated at a minimum frame rate of 20 Hz by STISIM

Drive version 2.08.02 (Systems Technology, Inc., Hawthorne, CA) in response to the subject's driving inputs, including steering wheel turning, brake and accelerator pedal use. The vehicle's sound system played a set of standard driving noises, which consisted of accelerating and braking sounds, other vehicle noises, and engine feedback. A video camera was mounted on the dashboard directly behind the steering wheel to monitor eye movement and general behavior of the subject. A second camera mounted above the driver's shoulder recorded interactions with the navigation systems. Both cameras recorded data at 30 frames per second.

As illustrated in Figure 1, two types of navigation systems were employed: a dashboard-mounted portable Garmin GPS unit (nüvi 1490LMT 5-inch (12.7 cm) backlit TFT color touchscreen display with 480x272 WQVGA pixel resolution supplied with City Navigator NT data) and an iPod Touch "Google Maps" application (identical to the Apple iPhone iOS 5 Google Maps application). The iPod was initially placed in the vehicle cup holder, though participants were allowed to use and reposition the iPod as they normally would throughout the simulated drive.

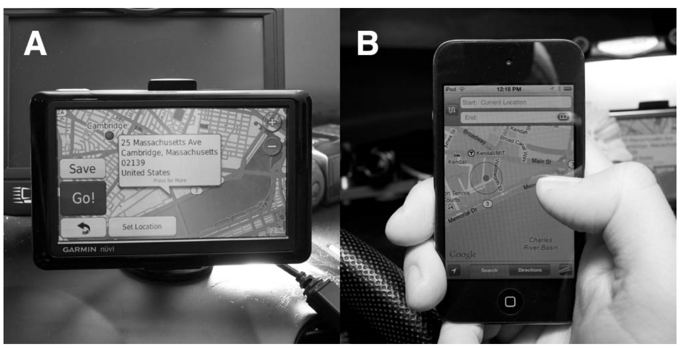

Figure 1. Photos of the Garmin (A) and iPod Touch (B) devices. Addresses to be entered into the navigation system were presented on index cards (not pictured) mounted on top of the dashboard just behind the steering wheel.

Procedure

Participants read and signed an informed consent form, eligibility was verified by interview, and primary and supplemental questionnaires that included items related to cell phone and GPS experience, type usage, and preferences were completed. A brief explanation of the experimental protocol was given and the subject was instructed to enter the simulator and adjust the driver's seat as desired.

Recorded audio instructions described the simulator and provided the following guidance and incentive: "During the study, you will receive a monetary award for performing the tasks while you continue driving the simulator. While performance on the tasks is important, you should balance driving safety while you attempt to complete the tasks, just as you would when driving a real car. Since in the real world you cannot disregard the traffic code, you may be penalized $2 for every ticket you receive and $5 for any collision." These instructions were intended to encourage participants to realistically balance the experiment tasks with safe driving (Reimer et al., 2006). In actuality, all participants received equal compensation.

The simulation scenario consisted of a divided highway with two lanes in each direction and a 2 ft (0.61 m) shoulder on each side of the roadway. Lane width was 15 ft (3.62 m) and the posted speed limit for the evaluation portions of the drive was 65 mph (104.6 km/h). Typical traffic events on the virtual highway included passing vehicles, lane changes, and slow downs. The average traffic density in the virtual scenario was set at 23 vehicles/mile (14.3/km). Average traffic speed for vehicles in the left lane was set equal to the posted speed limit of 65 mph (104.6 km/h) and 5 mph slower (96.5 km/h) for the right lane.

The experiment was divided into three blocks: familiarity drive, GPS task, and iPod task. The familiarity drive allowed the subject to practice controlling and responding to the simulator during about five minutes of uninterrupted driving. The order of the GPS and iPod evaluation blocks was counterbalanced across subjects.

The two evaluation blocks began with the subject parked. Subjects were walked through a practice entry and then instructed to enter a second practice address without guidance. Additional practice entries were offered until the subject was comfortable with his or her ability to complete the address entry task.

Once the subject was sufficiently trained on destination entry with the appropriate device, the research associate started the simulation. While driving, the subject was prompted to enter two addresses. The addresses were presented on index cards placed on the dashboard behind the steering wheel. The addresses all consisted of 3 digit street numbers, a street name and type, city name, and both the state and zip code were provided. When a subject had fully completed two entries, they were instructed to stop the vehicle. The research associate then presented a clipboard with two workload rating worksheets: a global workload rating scale, 0 (low) – 10 (high) and the NASA TLX. The subject was instructed to provide both workload ratings for the device task just completed.

Data Reduction and Analysis

Eye data were processed following ISO standards (ISO 15007-1, 2002; ISO 15007-2, 2001) and counts of glances greater than 2 seconds were computed. The percentage of time spent with eyes on the device was computed as a ratio of glance duration to the device by the total length of a task. Mean vehicle velocity and the root mean square of lane deviation were also computed for each device use period as well as a baseline single task driving reference period immediately prior to initiation of the address entry task.

Owing to the relatively small sample size and non-normality present in several relevant sub-samples, primary statistical tests were conducted using nonparametric alternatives, such as the Wilcoxon signed rank test and Friedman's test.

RESULTS

Vehicle telemetry data are summarized in Table 1. There were no significant differences in either mean velocity or deviation of lane position across baseline and task periods (velocity $X^2_{(2)}$ = 3.0, p = .223; lane deviation $X^2_{(2)}$ =.083, p = .959)

Table 1. Vehicle telemetry measures for mean velocity and root mean square lane deviation during each block. Standard deviations in parentheses.

Block	Velocity (mph)	Lane Deviation (ft)
Baseline	63.5 (2.4)	1.49 (0.62)
iPod	62.6 (2.2)	1.50 (0.54)
Garmin	62.4 (2.1)	1.41 (0.50)

Eye glance metrics were computed based on the amount of time spent looking off road (consistent with NHTSA guidelines), and alternatively, based on the amount of time spent glancing toward the task interface. The 2012 NHTSA visual-manual distraction guidelines recommend that for a task to be considered acceptably safe, 85% of tested subjects must: a) spend less than 12 seconds glancing off the road, b) maintain a mean off road glance time of no more than 2 seconds, and c) complete the task with a long glance rate (glances of duration greater than 2 seconds) of less than 15%. In our sample of 23 subjects, this means that a minimum of 20 subjects must pass each metric for the task to meet NHTSA's guidelines. (Note: NHTSA's guidelines recommend a sample in which subjects are equally distributed across the age groups of 18-24, 25-39, 40-54, and 55+. Since older subjects are expected to have more difficulty with device interaction, failure to meet the guidelines in a younger sample such as this would definitely constitute a failure, while a narrow "pass" of the criteria could not be assumed to indicate that the device would pass testing that includes the full recommended age distribution.) Subject performance on all three NHTSA criteria is shown in Figures 2-4. Frequency of glances is summarized in Figure 5.

For the off road glance metrics, the effect of device (iPod application or Garmin navigation system) on total glance time was not significant (29.4s Garmin, 29.8s iPod, V = 131, p = .842, paired Wilcoxon signed rank test). Mean glance time differed significantly by device (0.98s Garmin, 1.11s iPod, V = 229, p = .004), as did percent of long glances (6.9% Garmin, 9.8% iPod, V = 145, p = .046). Total number of glances was significantly higher for the Garmin system (30.4) as compared to the iPod (27.1) (V = 60.5, p = .019).

The same pattern of results appears when only glances to the device are considered. Total glance time was not significantly affected by device type (23.6s Garmin, 24.8s iPod, V = 150, p = .731). Mean glance time (1.1s Garmin, 1.29s iPod), percent of long glances (9.6% Garmin, 13.5% iPod), and total number of glances (21.7 Garmin, 19.2 iPod) were significantly different between device types (V = 237, p = .002; V = 148, p = .035; V = 68.5, p = .036, respectively).

A comparison of measurement methods (off road glance time vs. device glance time) shows a main effect of measurement type on all metrics. Total glance time was 5.1s longer when glances were measured based on total time off road as opposed to on the task device (V = 84, p < .001). There were also 7.9 more glances (averaged across both devices) during the task under the off road measurement method (V = 20.5, p < .001). Mean glance time was 0.13s lower for the off road measurement condition (V = 934, p < .001). The percentage of glances greater than 2.0s was slightly greater in the device measurement condition (9.7% vs. 7.1%, V = 541, p = .004).

Workload ratings for the Garmin and iPod tasks were not significantly different on either scale (TLX, V = 136, p = .964; Global, V = 130.5, p = .612).

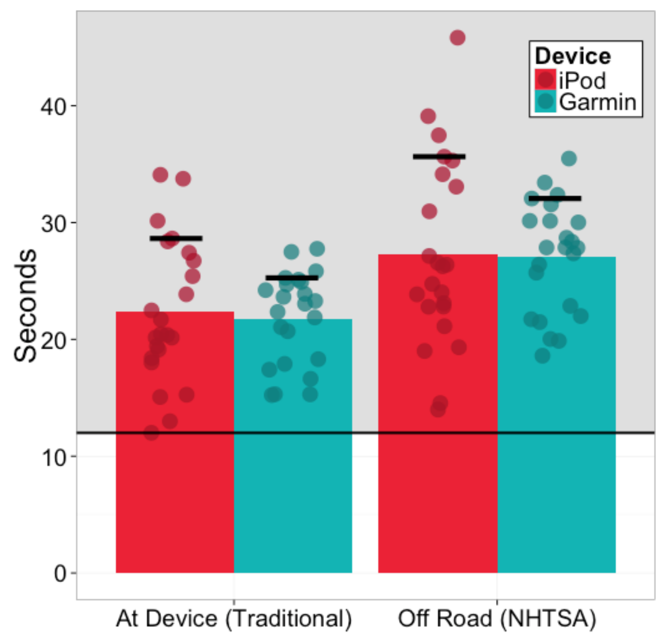

Figure 2. Total glance time. Bars represent sample means. Horizontal line segments represent the 85th percentile performance for each sample. Dots represent individual subjects. Dot positions have been jittered horizontally and made partially transparent to minimize overplotting of data points. Horizontal line represents NHTSA guideline criterion value of 12 seconds.

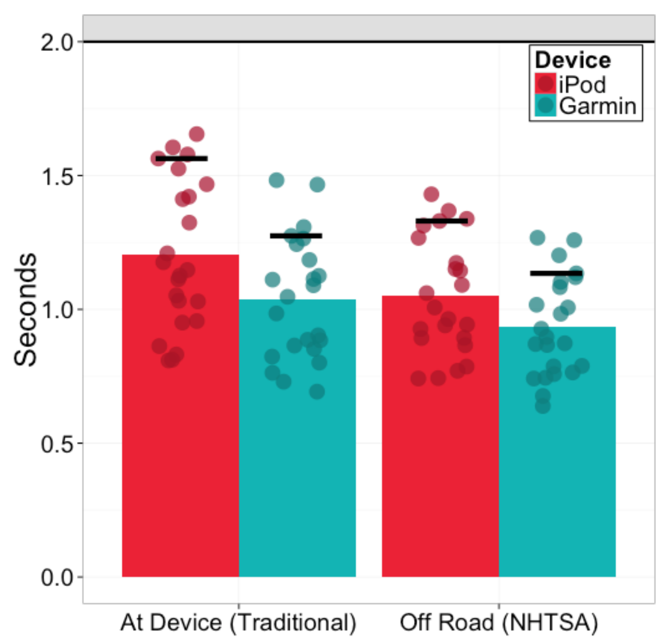

Figure 3. Mean glance duration. Horizontal line near top of graph represents NHTSA guideline criterion value of 2 seconds.

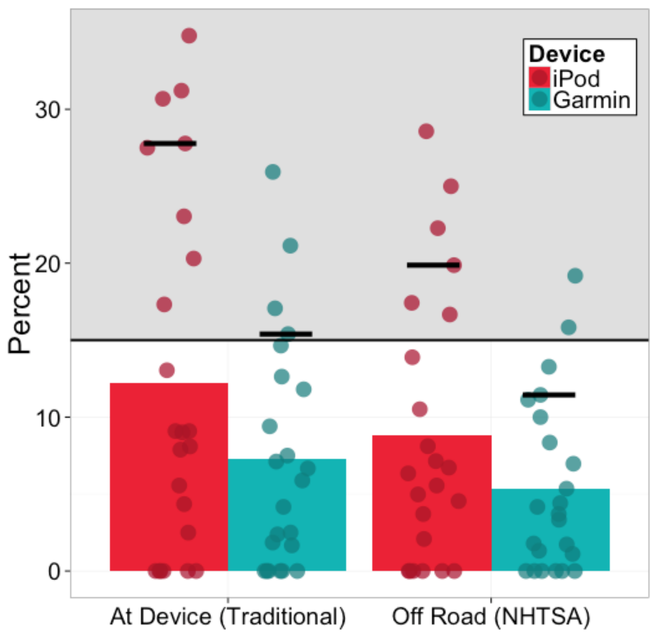

Figure 4. Percent of glances greater than 2.0 seconds. The horizontal line represents NHTSA guideline criterion value of 15 percent.

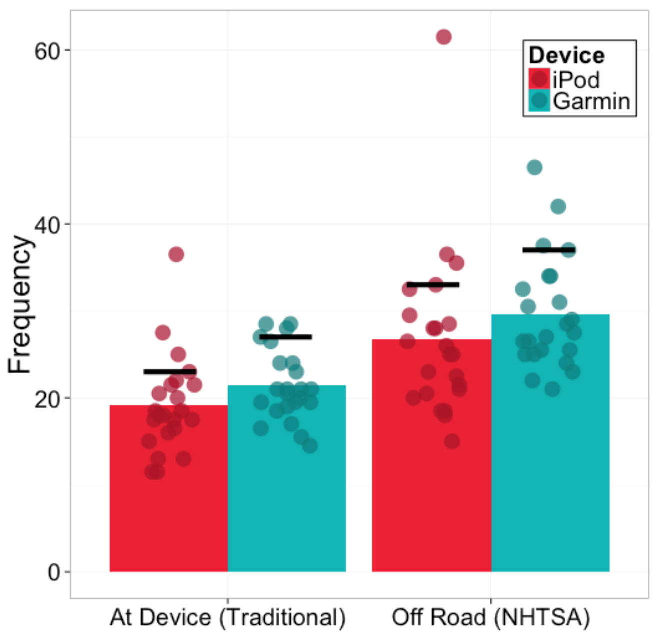

Figure 5 Total number of glances.

Table 2 summarizes the percentage of subjects who passed the NHTSA distraction criteria. For off road glance metrics, no subjects were able to complete the navigation entry task in less than 12 seconds of total glance time, regardless of the device used. Conversely, all subjects maintained an average glance time of less than 2 seconds. An acceptable percentage of the sample was able to maintain a long glance rate of less than 15% when the task was performed with the Garmin device, but not with the iPod.

For the device glance metrics, no subjects completed the task with a total glance time of less than 12 seconds, regardless of the device used. However, all subjects maintained an average glance time of less than 2 seconds. Finally, an unacceptably low percentage of subjects maintained a long glance rate of 15% or less (82.6% for the Garmin, 65.2% for the iPod).

Table 2. Percent of sample meeting the NHTSA distraction criteria cutoffs. NHTSA's guidelines state that 85% of the sample must meet these criteria for the investigated task to be considered safe. Metrics failing to meet this cutoff are highlighted in red.

Device	Looking	Total Time	Mean Glance Time	< 15% 2s Glances
iPod	At Device	**0%**	100%	**65.2%**
	Off Road	**0%**	100%	**73.9%**
Garmin	At Device	**0%**	100%	**82.6%**
	Off Road	**0%**	100%	91.3%

Note: The Off Road measure corresponds to the NHTSA guidelines.

DISCUSSION

The present study reveals several interesting findings. Although participants were able to complete the navigation entry task in the same amount of total glance time regardless of the device used, the device appears to impact participants' visual *strategies*. Use of the Garmin system, as opposed to the iPod, resulted in a larger number of off road glances, significantly shorter mean glance time, and a significantly lower percentage of long glances. This may indicate that drivers are able to shift visual attention to and from secondary tasks more quickly if the interface is mounted vs. being hand-held. It can also be noted that the Garmin system divides address entry into a number of discrete steps, whereas the iPod application integrates the process into a single, continuous action. As a result, it may have been more natural to allocate individual glance durations more in accordance with the NHTSA guidelines concerning percentage of glances < 2 seconds with the discrete steps of the Garmin system.

If the NHTSA criteria for embedded visual-manual interfaces are applied to these portable devices and applications, performing a manual navigation entry task with a dash-mounted Garmin system or iPod application would likely not pass, as none of our participants were able to complete the navigation entry task within NHTSA's recommended total off road glance time of 12 seconds or less. On the other hand, all participants maintained a mean glance time of less than 2 seconds. This indicates that while drivers do a fairly good job according to the guidelines of allocating individual glances between the road and the secondary task, it seems that the task itself forces the driver to take his/her eyes off the road for what is considered an unacceptably lengthy total time. If navigation entry is to be undertaken while underway, it is clear that interfaces will need to be developed that reduce the total time that a driver must take their eyes off the road and minimize long duration glances.

NHTSA recommends that any time glancing off road be counted when computing distraction metrics. However, not all glances off the forward roadway are necessarily related to a distracting activity, e.g. when a driver checks side mirrors to safely perform a lane change. As illustrated here, the metrics could alternatively be computed based only on glances to the navigation device, essentially excluding glances to the mirrors, speedometer, and glances towards a cue card used to present the driver with an address for entry. Taken together, the data suggest that device-relevant glances are longer than other non-road glances. Therefore, including all non-road glances has the potential to mask some of the true visual demands of secondary task performance. A perhaps unintended consequence of including non-device oriented glances in NHTSA's total off road glance assessment approach is that the mean glance duration and percentage of glances less than 2 seconds metrics are both lower than is the case when the glances to device approach is used. Such lower values might result in some testing situations in a device that would not have met the Alliance criteria to pass the new NHTSA guidelines. Weighing the drawbacks and benefits of each approach is a complex topic area that deserves continued investigation and discussion. This observation is in line with NHTSA's explicit comments on the need for further research.

Another methodological consideration seems worthy of mention. Procedures designed to support task presentation, such as an address cue card, may impact the overall rating of a device by adding to off road glance time. Conversely, requiring a subject to memorize addresses or phone numbers to avoid adding visual demand may increase cognitive load and have unintended consequences. The implications of both approaches for the type of activities that can be effectively assessed is a topic for future research.

In summary, the NHTSA performance metrics developed for use with embedded visual-manual interfaces were applied to portable aftermarket devices and applications – demonstrating their potential utility beyond the scope of the Phase 1 NHTSA guidelines. However, the change of computational method from device related glances to total off road glances can impact the proportion of the experiment sample that passes the NHTSA safety criteria. A number of methodological considerations need to be revisited in developing safety guidelines for emerging technologies. As noted earlier, in interpreting our results it must be taken into account that two subjects were excluded from the analysis due to task completion times and/or glance frequencies that were more than three standard deviations from the group mean. Therefore, the results, as presented, may represent a lower-bound estimation of the visual demand of the devices.

In looking at ways to improve this research, confidence in eye glance data integrity could be enhanced through dual coding with mediation. Future work might expand upon the self-report data to explicitly evaluate touchscreen vs. non-touchscreen and smartphone vs. non-smartphone users, compare fixed vs. non-fixed devices to isolate the effects of holding a device, and expand the sample size and age distribution to meet the NHTSA criteria.

ACKNOWLEDGEMENTS

Support for various aspects of this research was provided by Denso Automotive, the U.S. DOT's Region One New England University Transportation Center, and the Santos Family Foundation. We gratefully acknowledge Linda Angell for a thoughtful review and comments, Alea Mehler and Hale McAnulty for subject recruitment, protocol development, data collection, and other activities and Yan Yang for training the first 3 authors in eye glance coding.

REFERENCES

Basacik, D., Reed, N., & Robbins, R. (2011). Smartphone use while driving: a simulator study. Transport Research Laboratory Report PPR592, Berkshire, UK..

Driver focus-telematics working group of Alliance of Automobile Manufacturers (AAM), 2006. "Statement of principles, criteria and verification procedures on driver interactions with advanced in- vehicle information and communication systems." http://www.umich.edu/~driving/safety/guidelines.html (accessed 12/5/2012)

Jensen, B. S., Skov, M. B., Thiruravichandran, N. (2012). Studying driver attention and behaviour for three configurations of GPS navigation in real traffic driving. *Proceedings of CHI 2010: Driving, Interrupted*, Atlanta, GA, USA, April 10-15, 2010.

Lee, W., & Cheng, B. (2010). Comparison of portable and onboard navigation system for the effects in real driving. *Safety Science, 48(10),* 1421-1426.

NHTSA (2010). *Overview of the National Highway Traffic Safety Administration's Driver Distraction Program.* National Highway Transportation Safety Administration. (No. DOT HS 811 299).

NHTSA (2012). Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices. Department of Transportation. National Highway Traffic Safety Administration. Docket No. NHTSA-2010-0053. 2010.

Oracle (2011). Opportunity calling: the future of mobile communications - take two. Oracle communications. http://www.oracle.com/us/corporate/press/522729 Visited: December 10, 2012.

Quaresma, M. (2012). Assessment of visual demand of typical data entry tasks in automotive navigation systems for iphone. *Work, 41,* 6139-6144.

Reimer, B., D'Ambrosio, L. A., Coughlin, J. F., Kafrissen, M. E., & Biederman, J. (2006). Using self-reported data to assess the validity of driving simulation data. *Behavior Research Methods*, *38(2),* 314–324.

Reimer, B., Mehler, B., Donmez, B., Pala, S., Wang, Y., Olson, K., et al., (2012). A driving simulator study examining phone dialing with an iPhone vs. a button style flip- phone. *Proceedings of the 56th Annual Meeting of the Human Factors and Ergonomics Society*, Boston, MA, 2012, pp. 2191-2195.

Samuel, S., Pollatsek, A., & Fisher, D. L. (2011). Texting while driving: Evaluation of glance distributions for frequent/infrequent texters and keypad/touchpad texters. *Proceedings of the 6th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Lake Tahoe, CA, 2011, pp. 424-432.

Tijerina, L., Parmer, E., Goodman, M. J. (1998). Driver workload assessment of route guidance system destination entry while driving: a test track study. *Proceedings of the 5th ITS World Conference*, Seoul, Korea, October 12-16, 1998. (Compact Disk)