A Simulation Study Examining Smartphone Destination Entry while Driving

Daniel Munger, Bruce Mehler, Bryan Reimer*, Jonathan Dobres,
Anthony Pettinato, Brahmi Pugh & Joseph F. Coughlin
MIT AgeLab & New England Univ. Transportation Ctr.
77 Massachusetts Avenue, E40-291
Cambridge, MA 02139
*Corresponding author: reimer@mit.edu - (617) 452-2177

ABSTRACT
A driving simulation study was performed to compare visual-manual (touch screen based) destination entry using a Samsung Galaxy S4 smartphone with the standard voice command based interface and a voice based "Hands-Free mode" that appears to be intended for use while driving (i.e. has a steering wheel icon adjacent to the mode selection menu and the voice interface menu screen is visually austere when compared with the standard voice mode). The performance of 24 drivers on an alphanumeric street address entry task was assessed with respect to subjective workload, task duration, standard deviation of lateral lane position, response to a detection response task (DRT), and heart rate. With the exception of heart rate, all evaluation measures indicate that the voice interfaces provide significant advantages over the touch interface. Furthermore, subjective workload ratings and task duration measures imply that the "Hands-Free" voice based mode may have some costs relative to the standard voice command based interface. Lastly, all destination entry methods were associated with an increased DRT reaction time and higher miss-rates compared to a baseline driving condition.

Author Keywords
Automotive human machine interface; Voice interface; Speech interface; Distraction; Driver safety; Detection response task (DRT); Workload.

ACM Classification Keywords
H.1.2 **[User/Machine Systems]**: Human Factors; Human Information Processing; H.5.2 **[User Interfaces]**: Ergonomics; Interaction styles; User-Centered Design; Voice I/O

AutomotiveUI '14, September 17 - 19 2014, Seattle, WA, USA

ACM 978-1-4503-3212-5/14/09…$15.00
http://dx.doi.org/10.1145/2667317.2667349

INTRODUCTION
Despite the proliferation of embedded vehicle navigation systems, many drivers prefer to utilize the navigation applications native to their smartphones. In 2012, 47% of drivers reported using a smartphone based navigation application, up from 37% the previous year [4]. Also in 2012, 65% of smartphone owners reported using the turn-by-turn navigation features on their phones while driving [9]. As smartphone use continues to climb, it is becoming increasingly important to understand how these devices can be integrated into the driving environment while minimizing the attentional demand drawn away from the driving task. Many smartphone applications can be utilized, in part, through voice-operated features with the aim of reducing visual attentional demands. However, it is not clear how these multi-modal interaction modes impact workload.

It has been argued that voice command based systems are inherently less demanding, and presumably safer, than visual-manual interfaces [7][13]. Although there is strong data supporting this argument, studies of more complex voice initiated interactions show significant visual demands and highlight the need for more comprehensive examinations of the multi-modal demands placed upon the driver [5][11].

While NHTSA has issued guidelines regarding the demands of embedded visual-manual interfaces [6], research comparing drivers' operation of portable electronics using different modalities has been limited (see [15] for one such study). In the current study, we focus on investigating drivers' use of different interface modes for destination entry into the Google Maps Application on a Samsung Galaxy S4 smartphone. This smartphone model was selected because it is widely used and is representative of current voice-operated and visual-manual features on commercially-available smartphones.

METHODS

Participants
A total of 30 drivers were recruited across two age groups, 20–24 and 55+. Participants were required to be experienced drivers, as determined by having held a valid

driving license for 3+ years and having a self-reported driving frequency of at least once per week. Additional requirements included: being in self-reported reasonably good health for one's age, being fully comfortable speaking and reading English, and having no major illness resulting in hospitalization in the past 6 months. A diagnosis of Parkinson's disease or other neurological problems was also an exclusion criterion due to the possible impact on fine motor control. The study was approved by the Committee on the Use of Humans as Experiment Subjects (COUHES) of the Massachusetts Institute of Technology and compensation of $40 was provided for participation.

Apparatus

The study utilized the MIT AgeLab fixed based driving simulator, a full cab Volkswagen New Beetle with a front projection system providing a view of approximately 40 degrees. Graphical updates were generated using STISIM Drive version 2.08.02 (Systems Technology, Inc., Hawthorne, CA) based upon a driver's interaction with the steering wheel, brake, and accelerator. Instructions and audio tasks were pre-recorded and presented through the vehicle sound system. Correspondence between the demands of this simulator configuration and actual driving scenarios has been established through previous research (see [10] for one such study). Measures of heart rate (modified lead II EKG configuration) and electrodermal activity (skin conductance) were recorded at 250 Hz using a MEDAC System/3 physiological monitoring unit. The driving scenario consisted of a two lane rural road with a posted speed limit of 50 mph.

A CogLens remote mounted Detection-Response Task (DRT) was implemented in accordance with the draft ISO Standard [3], which advocates the DRT as a measure of cognitive distraction while driving. A red LED was mounted on the windshield near the center of the participant's field-of-view and responses were recorded from a micro-switch placed on the participant's left index finger. The LED was activated every 3–5 seconds for a period of one second or until the participant responded using the finger-mounted switch.

Destination Entry Device

Participants entered destination addresses in a Samsung Galaxy S4, model number SCH-1545 (released March 2013), which featured a 5" display with 1920 x 1080 resolution. The mobile network carrier was Verizon Wireless and the operating system was Android 4.3. The device was "free floating" (not mounted) and participants held it in their hand or rested it on the center console or other location at their discretion while performing tasks.

Navigation tasks were carried-out using the Google Maps application, and participants were extensively trained on how to enter an address using the three different interaction modes outlined in Table 1. One method required visual-manual touchscreen interaction and two were auditory-vocal-visual-manual (e.g. voice based commands).

For the touch interface, participants opened the Google Maps application, typed a specified address into the search bar, selected a car icon to show driving routes, selected a route and tapped "Start navigation." In the "Hands-Free" mode, participants enabled the voice recognition feature by double pressing the phone's home button. The phone then presented a seemingly random introduction, e.g. "Hello, I hope you are having a great and productive week. If you need any help say 'Hi Galaxy'". After saying "Hi Galaxy", the participant would then say "navigate to," followed by the street address. In the "standard" voice mode, the verbose audio introduction and "Hi Galaxy" command were omitted; after double-tapping the voice button, participants would immediately speak the navigation command.

Procedure

Participants first read and signed an informed consent and eligibility was confirmed by a verbal interview. Participants were instructed on how to perform the navigation tasks, and were given an opportunity to practice entering an address for all three destination entry modes while seated in the lab. Once participants were able to correctly enter an address using all three modes, they moved to the driving simulator where they learned how to perform the DRT, completed an introductory drive, and then practiced the dual task of driving and responding to the DRT.

The experimental period consisted of three counterbalanced blocks corresponding to each of the destination entry modes. Following the draft ISO guidelines, each experiment block began with a training period building-up to the triple task of driving, responding to the DRT and entering a destination. Participants performed the destination entry task while stationary, first without and then with the DRT. The process was then repeated while driving. Participants were required to achieve proficiency on each training condition (defined as performing the task correctly while also responding to at least 70% of the DRT stimuli) before advancing to the next stage.

Step	Voice: Standard	Voice: Hands-Free	Manual (Touchscreen)
1	Tap Home button to wake up screen		
2	Double-tap Home button		Open Google Maps application
3	Speak: "Navigate to 3-8-5 Prospect St, Cambridge"	Speak: "Hi Galaxy"	Tap Search bar
4		Speak: "Navigate to 3-8-5 Prospect St, Cambridge"	Type address: "385 Prospect St, Cambridge" and select address when it appears
5			Select the car icon to show routes
6			Select "Start navigation"

Table 1. Destination entry steps for all three modes

During the evaluation periods, participants first completed 3 minutes of single-task driving and were then asked to enter the address "177 Massachusetts Ave, Cambridge" while simultaneously responding to the DRT. The task was considered complete once the participant stated that the device had finished calculating directions. Five seconds after completing the task, participants were asked to cancel the address by pressing the phone's "Back" button until they reached the home screen. Participants then experienced a 30 second separation period, followed by a 60 second baseline DRT period, followed by another 30 second separation. They then completed a second destination entry ("293 Beacon St, Boston") while responding to the DRT and then again returned to the home screen. Once the block was complete, participants were asked to stop the car and complete a global workload rating (0 – 10; see [11] for details on this scale) and the NASA-TLX scale [2]. These steps were repeated for all 3 destination entry modes.

Data Reduction and Analysis

Six participants were excluded from analysis, leaving a final sample of 24 participants equally balanced by gender and across the two age groups. Two participants were omitted due to EKG recording issues; two due to an inability to perform the task after five attempts; one participant was omitted because of a persistent server error reported by the smartphone that made it impossible to use the navigation features; and one participant was withdrawn due to simulator-induced nausea. Skin conductance signals showed evidence of significant movement artifact due to holding the phone / wheel with one hand and were not used in this analysis.

Values for task duration, standard deviation of lateral lane position, DRT reaction time and heart rate were computed as the mean over the two destination entry periods for each method. The baseline "just drive" period was computed as the mean across all three task blocks. DRT reaction times greater than 2500ms were counted as a "miss" in accordance with the draft ISO standard.

Error bars indicate mean-adjusted standard error. Statistical significance was determined by Friedman's test (for differences across conditions) and Wilcoxon signed rank test (for post-hoc pair-wise comparisons). These non-parametric tests are robust to the effects of outliers, non-normality, and the type of ordinal data produced from the workload scales. Analysis was performed using R [8]. Post-hoc test significance is reported when $p < 0.05$.

RESULTS

Subjective workload ratings are shown in Figure 1. Both rating scales show a significant main effect of mode (global $\chi^2(2) = 29.70$, $p < 0.001$; NASA-TLX $\chi^2(2) = 11.64$, $p = 0.003$). Post-hoc tests show significant differences between all three destination entry modes for the global rating scale, and between all modes except touch and Hands-Free voice for the NASA-TLX scale. Global workload and NASA-TLX scores did not significantly differ in any of the 3 modes (all $p > 0.24$, Wilcoxon test).

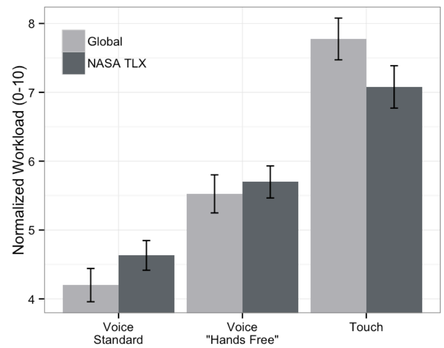

Figure 1. Subjective workload measures for all three destination entry modes.

As illustrated in Figure 2, a main effect of task duration appears ($\chi^2(2) = 24.25$, $p < 0.001$). Post-hoc tests indicate significant differences between the standard voice and Hands-Free voice mode, as well as between standard voice and touch ($p < .001$ for both).

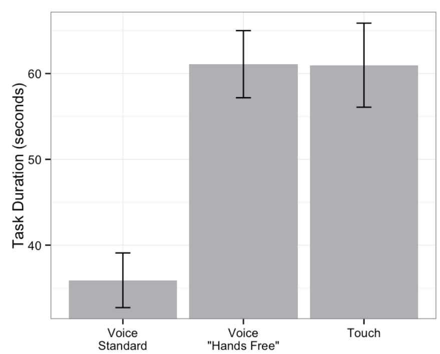

Figure 2. Task duration in seconds for all three destination entry modes.

Figure 3 shows standard deviation of lateral lane position. A significant main effect of condition appears ($\chi^2(3) = 30.20$, $p < 0.001$). Post-hoc tests show a significant difference between touch destination entry and the other three conditions, but reveal no statistical difference between Hands-Free voice, standard voice and the just drive condition.

As illustrated in Figure 4, DRT reaction time is impacted by condition ($\chi^2(3) = 51.95$, $p < 0.001$). Post-hoc tests show significance across all comparisons ($p < .01$ for all tests) except between voice modes. DRT error rate (not shown) demonstrates significant differences between all modes relative to single-task driving ($\chi^2(3) = 12.16$, $p = 0.007$), but post-hoc tests show no differences among the three destination entry modes. The DRT miss rates for the four conditions are as follows: just drive (M=0.98%, SE=1.35%), standard voice (M=9.19%, SE=1.65%), Hands-Free voice (M=8.33%, SE=0.94%), and touch entry (M=11.25%, SE=1.69%).

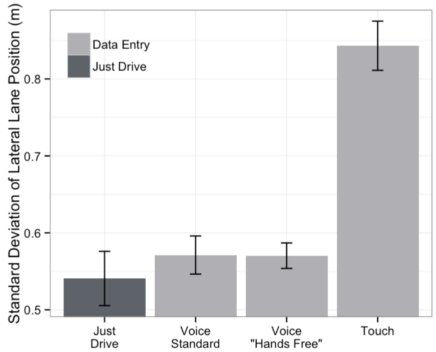

Figure 3. Standard deviation of lateral lane position.

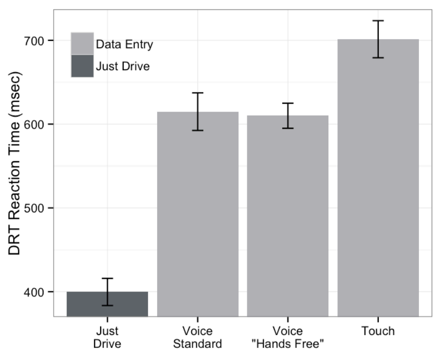

Figure 4. DRT reaction times in milliseconds

Finally, heart rate is displayed in Figure 5. A significant effect of condition appears ($\chi^2(3) = 37.70$, $p < 0.001$). Post-hoc tests show no significant difference between the three modes, but heart rate was significantly elevated compared with single-task driving for all three entry methods.

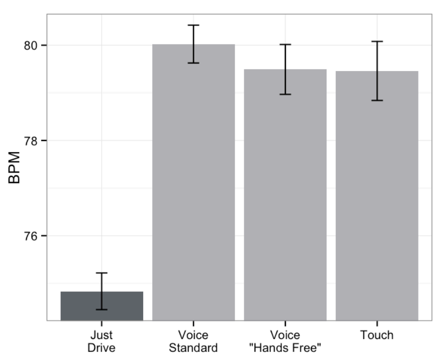

Figure 5. Heart rate in beats per minute.

DISCUSSION

These results reveal a number of important differences between the relative demands of the voice and touch interfaces studied. In particular, the voice based interfaces are associated with lower subjective workload, shorter task duration, less standard deviation of lateral lane position, and faster DRT response times. As shown in Figures 3 & 4, participants experienced significantly greater variability in lateral vehicle control and longer event detection response times while engaged in the visual-manual touch destination entry task relative to both of the voice destination entry methods. These results are consistent with prior findings showing increases in workload [14] and DRT reaction time ([1][12]) associated with touch as compared with voice interfaces.

The differences between the two voice modes are less apparent, as there are no distinctions between the interfaces with respect to DRT reaction times or standard deviation of lateral lane position. However, based on both subjective workload ratings, the Hands-Free voice mode was perceived as more difficult to use than the standard voice interface. Additionally, destination entry using the Hands-Free mode required an average of 25.18 seconds longer to complete compared to the standard mode. The relative difference in task duration might be traceable to several factors such as the auditory introduction, extra verbal prompting and confusion related to the additional layer of commands required to enter a destination. The duration of the Hands-Free task (M=59.92s, SE=3.30s) did not differ significantly from the touch based interface (M=59.98s, SE=4.27s).

Lastly, while heart rate and DRT miss rate do not reveal any significant differences between the three modes of entry, these results show significant differences between destination entry and the single-task driving condition. This implies that destination entry causes an increase in physiological workload and a reduction in event detection regardless of the interaction method.

Based upon these findings, it is clear that caution should be observed in the development and use of any in-vehicle voice interface. The Hands-Free voice mode does not appear to provide any advantages in heart rate, DRT or driving performance metrics over the standard voice interface, and the longer task duration may frustrate drivers, reduce utilization, and extend the period of increased workload. These results support earlier findings ([1][11][14]) that although voice interfaces are generally less distracting than visual-manual interfaces, some voice interfaces can be further optimized to reduce driver demand.

Limitations

The DRT response switch was placed on participants' left index finger in accordance with the draft ISO standard. This placement may have impacted participants' ability to respond to the DRT task while holding the smartphone. This suggests a possible limitation of the proposed DRT standard as a general method of measuring performance. One may hypothesize that DRT responses during operations that require manipulation involving one of the driver's hands may be impacted more than DRT responses to tasks that have limited or no hand involvement. Additionally, it is important to note that these findings measure demands of navigation destination entry tasks with respect to the simulation environment. While conceptual concerns related to the demands observed are likely relevant to understanding demand under actual driving conditions, they should be interpreted cautiously.

ACKNOWLEDGEMENTS

Support for this work was provided by 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. We also thank Adrian Rumpold, Mauricio Muñoz and Alexander Waldmann for their assistance.

REFERENCES

1. Conti, A.S., Dlugosch, C., Schwarz, F. and Bengler, K. 2013. Driving and Speaking: Revelations by the Head-Mounted Detection Response Task. *Proceedings of the 7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design.* (Bolton Landing, NY, Jun. 17-20, 2013). 362-368.
2. Hart, S.G. 2006. NASA-task load index (NASA-TLX); 20 years later. *Proceedings of the 50th Human Factors and Ergonomics Society Annual Meeting*, (Santa Monica, CA, Oct. 16-20, 2006). 904–908.
3. ISO 17488 (draft). 2013. *Road Vehicles - Transport information and control systems - Detection Response Task*: (New Work Item for ISO TC 22 SC13 WG8). International Standards Organization, Geneva, Switzerland, 2013.
4. JD Power McGraw Hill Financial. 2013. *2012 U.S. Navigation Usage and Satisfaction Study Results* (Jan. 2013), Retrieved April 22, 2014: http://autos.jdpower.com/content/study-auto/l3KYlac/2012-u-s-navigation-usage-and-satisfaction-study-results.htm
5. Mehler, B., Reimer, B., Dobres, J., McAnulty, H., Mehler, A., Munger, D., and Coughlin, J.F., *Further Evaluation of the Effects of a Production Level "Voice-Command" Interface on Driver Behavior: Replication and a Consideration of the Significance of Training Method (Technical Report 2014-2)*. MIT AgeLab, Cambridge, MA, 2014.
6. National Highway Traffic Safety Administration, *Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices (Docket No. NHTSA-2010-0053)*. U.S. Department of Transportation National Highway Traffic Safety Administration (NHTSA), , Washington, DC, 2013.
7. Owens, J., McLaughlin, S. and Sudweeks, J. 2010. On-road comparison of driving performance measures when using handheld and voice-control interfaces for mobile phones and portable music players, *SAE Int. J. Passeng. Cars - Mech. Syst.,* 3, 1 (Apr. 2010), 734-743.
8. R Core Team, *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.
9. Rainie, L. and Fox, S. 2012. *Just-in-time Information through Mobile Connections.* Pew Research Center (May 2012). Retrieved April 22, 2014: http://www.pewinternet.org/2012/05/07/main-report-16/
10. Reimer, B. and Mehler, B. 2011. The impact of cognitive workload on physiological arousal in young adult drivers: a field study and simulation validation. *Ergonomics* 54, 10 (Oct. 2011), 932-942.
11. Reimer, B., Mehler, B., Dobres, J. and Coughlin, J.F., *The Effects of a Production Level "Voice-Command" Interface on Driver Behavior: Reported Workload, Physiology, Visual Attention, and Driving Performance (Technical Report 2013-17a)*. MIT AgeLab, Cambridge, MA, 2013.
12. Schreiner, S., Beckers, N., Bertrand, P., Munger, D., Dobres, J., Mehler, B. and Reimer, B. 2014. *A Comparison of the Demands of Destination Entry using Google Glass and the Samsung Galaxy S4* Proceedings of the 58th Annual Meeting of the Human Factors and Ergonomics Society. Chicago, IL, 2014.
13. Shutko, J., Mayer, K., Laansoo, E. and Tijerina, L. 2009. *Driver workload effects of cell phone, music player, and text messaging tasks with the Ford SYNC voice interface versus handheld visual-manual interfaces*, SAE Technical Paper 2009-01-0786.
14. Tsimhoni, O., Smith, D. and Green, P. 2004. Address Entry While Driving: Speech Recognition Versus a Touch-Screen Keyboard, *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 46, 4 (Jan. 2010), 600.
15. Yager, C.E. 2013. Driver Safety Impacts of Voice-to-Text Mobile Applications. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (San Diego, CA, Sep. 30 – Oct. 4, 2013) 1869-1873.