Proceedings of the Human Factors and Ergonomics Society Annual Meeting
http://pro.sagepub.com/

Comparing the Demands of Destination Entry using Google Glass and the Samsung Galaxy S4

Niek Beckers, Sam Schreiner, Pierre Bertrand, Bryan Reimer, Bruce Mehler, Daniel Munger and Jonathan Dobres

Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2014 58: 2156
DOI: 10.1177/1541931214581453

The online version of this article can be found at:
http://pro.sagepub.com/content/58/1/2156

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
Human Factors and Ergonomics Society

Additional services and information for *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* can be found at:

Email Alerts: http://pro.sagepub.com/cgi/alerts

Subscriptions: http://pro.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://pro.sagepub.com/content/58/1/2156.refs.html

>> Version of Record - Oct 17, 2014

What is This?

Downloaded from pro.sagepub.com by guest on October 18, 2014

Comparing the Demands of Destination Entry using Google Glass and the Samsung Galaxy S4

Niek Beckers[1], Sam Schreiner[1], Pierre Bertrand[1], Bryan Reimer[2*], Bruce Mehler[2], Daniel Munger[2], Jonathan Dobres[2]
[1]MIT Department of Aeronautics and Astronautics, Man Vehicle Lab
[2]MIT AgeLab and New England University Transportation Center
*Corresponding Author: reimer@mit.edu

A driving simulation study assessed the impact of vocally entering an alpha numeric destination into Google Glass relative to voice and touch-entry methods using a handheld Samsung Galaxy S4 smartphone. Driving performance (standard deviation of lateral lane position and longitudinal velocity) and reaction to a light detection response task (DRT) were recorded for a gender-balanced sample of 24 young adult drivers. Task completion time and subjective workload ratings were also measured. Using Google Glass for destination entry had a statistically higher miss rate than using the Samsung Galaxy S4 voice interface, the Google Glass method took less time to complete, and the two methods were given comparable workload ratings by participants. In agreement with previous work, both voice interfaces performed significantly better than touch entry; this was seen in workload ratings, task duration, lateral lane control, and DRT metrics. Finally, irrespective of device or modality, destination entry significantly decreased responsiveness to events in the forward scene (as measured by the DRT reaction time) as compared to the baseline driving.

INTRODUCTION

Although navigation devices started appearing in vehicles in the mid 1980's, it was not until recently that in-vehicle navigation devices have come into widespread use (Quaresma, 2012). This can be attributed to the introduction of navigation applications on smartphones coupled with the dramatic increase in smartphone use (Oracle, 2011; Quaresma, 2012) and an overall increase in the availability of embedded navigation systems. Navigation systems have been shown to create visual, auditory and cognitive distraction (Ranney, 2008; Dopart et al. 2013). Destination entry can occupy drivers for significant amounts of time and has been identified as the most demanding part of using a navigation device while driving (Young & Regan 2003).

Due to potential detrimental effects of mobile device use in vehicles, it is vital to study new devices before their introduction to the market to evaluate their impact on driving performance and safety. The recent unveiling of Google Glass as a new form of wearable mobile device presents a drastic departure from existing smartphones and other mobile devices; it is a voice-input, hands-free device that uses audio and visual feedback to interact with the user. While it has long been postulated that voice interfaces are less distracting than visual-manual interfaces (Tijerina, Parmer, & Goodman, 1998), recent research suggests that voice command interfaces are often multi-modal, calling upon a sizable level of visual resources (Reimer, Mehler, Dobres, & Coughlin, 2013). Therefore, it is an open question as to how drivers will allocate attention to the novel head-mounted screen as they use the vocal input and receive auditory and visual feedback. Furthermore, it is unclear how Google Glass affects driver behavior in comparison to voice-interface and touch-interface smartphone navigation applications.

These questions are timely. In October 2013, a California woman was given a citation for distracted driving due to the fact she was wearing Google Glass while driving (Reuters, 2014). The woman pled "Not Guilty" and the citation was dismissed by a California judge due to insufficient evidence that Google Glass was operating when the ticket was given. The bourgeoning national debate concerning the impact of Google Glass on driving performance and safety is a popular one, but lacking in scientific data. It seems likely that there will be a natural tendency to use Google Glass while driving, and its head-mounted nature may lead the general populace to assume that it is safer to use than traditional hand-held devices.

To this end, the authors conducted a study to analyze the impact of destination entry tasks using Google Glass on workload, driving performance, and attention. Google Glass was compared against two different benchmarks: voice and touch interfaces with the Samsung Galaxy S4 smartphone. The operation of Google Glass requires a wearer to learn a novel mental model of operation. Given our objective to understand the demands of the interfaces under study (as opposed of the ability of one to learn the interaction model of Google Glass), this study was framed by drawing the majority of participants from a population of highly educated students that would be expected to be a best case model of lead technology adopters.

METHODS

Participants

A total of 25 participants between the ages of 22 and 33 were recruited from the MIT campus and other nearby academic and research institutions. Participants were required to have held a valid driver's license for 3+ years and be in self-reported reasonably good health for their age. To reduce bias against voice interfaces, only native English speakers were considered. The research protocol was approved by MIT's IRB and compensation of $40 was provided.

One participant was excluded from analysis for being unable to successfully use the Samsung voice interface,

Copyright 2014 Human Factors and Ergonomics Society. DOI 10.1177/1541931214581453

Downloaded from pro.sagepub.com by guest on October 18, 2014

leaving a gender-balanced sample of 24 participants. The mean age of the analysis sample was 25.0 years (SD 2.6).

Apparatus

The driving simulator was a stationary, full cab 2001 Volkswagen Beetle situated in front of an 8' by 6' (2.44m by 1.83m) projection screen with approximately a 40 degree view of a virtual highway. Graphical updates were generated at a minimum frame rate of 20 Hz by STISIM Drive version 2.08.02 (Systems Technology, Inc.) in response to the participant's driving inputs, including steering wheel, brake and accelerator pedal use. The vehicle's sound system played a set of driving noises, which consisted of accelerating and braking sounds, other vehicle noises, and engine feedback.

The simulation scenario consisted of a two lane rural road with 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 during the evaluation period was 50 mph (80.5 km/h). The average oncoming traffic density was 9 vehicles/mile (5.6/km), staggered by +/- 200 ft (.061 km) to prevent uniform spacing. The speed of oncoming vehicles was equal to the posted speed limit of 50 mph. There were 8 straight segments, 2 left curves and 2 right curves, which were randomly shuffled for each drive. Curves became visible at a distance of 2000 ft (610 m) and had a curvature of 0.00015 (1/ft) (0.000492 1/m) and a length of 2000 ft (610 m).

A Detection Response Task (DRT) was implemented via a CogLens remote-mounted LED stimulus and a finger-mounted micro-switch response (http://coglens.com). In compliance with the draft ISO Standard (ISO 17488, 2013) for utilizing a remote visual stimulus, the LED was mounted near the center of the participant's field-of-view on the windshield. Following the proposed standard, the LED activated for 1 second with a uniformly distributed inter-stimulus interval between 3 and 5 seconds. The LED deactivated after 1 second or when the participant responded using the micro-switch. The micro-switch was placed on the participant's left index finger such that it could be pressed against the steering wheel while the participant's hand was in a neutral driving position.

Destination Entry Devices

Participants were evaluated on three different destination entry methods: voice entry using Google Glass XE11 and both voice and touch entry using the Samsung Galaxy S4 (see Figure 1). The Google Glass XE11 was released in February 2013 as an "Explorer" test program and features a head-mounted screen. The transparent screen is located above the line of sight of the right eye when looking straight ahead at the visual scene; to see the screen participants had to look up. A tactile interface is located on the right side of Google Glass. A voice interface is primarily used to interact with the device, during which the screen displays visual feedback. Google Glass required a connection to the Samsung Galaxy smartphone to enable navigation. To enter a destination into the Google Glass, participants tapped the tactile interface once to wake up the device and said "Ok Glass, get directions to 123 Central Avenue, Cambridge, Massachusetts." Participants were specifically instructed to say "one-two-three" rather than "one-hundred-twenty-three". After participants said "Ok Glass", the screen displayed a list of options for the participant to say, among which was "Get directions to…". During training participants became familiar enough with Google Glass to not rely on visual feedback from the screen. Participants wore Google Glass continuously on their head during the Google Glass evaluation period.

The Samsung Galaxy S4 is a commercially available smartphone released in March 2013. It is equipped with a 5-inch touch screen and voice interaction software called S-Voice. The smartphone was configured to use Google Maps for navigation for all input methods. To access the voice interface, participants were instructed to press the home button twice from the phone's home screen. The S-Voice "enhanced" voice interaction feature was disabled due to the variability in its audio greeting whenever the user attempted to use the voice interface. To enter an address using the Samsung Galaxy, participants tapped the home button once to wake up the device, and then twice to enter voice recognition mode. Participants then said "Navigate to 123 Central Avenue, Cambridge, Massachusetts" with the same numeric pronunciation instructions as Google Glass. To enter a destination using touch input, participants opened the Google Maps application, manually entered the destination in the dialog box, and then tapped the "get directions" button. While manually entering a destination, Google Maps showed a list of suggestions. Participants were allowed to use these suggestions after entering the address number and at least one letter of the street name. Participants were instructed to place the phone in or around the central console when not engaged in an entry task.

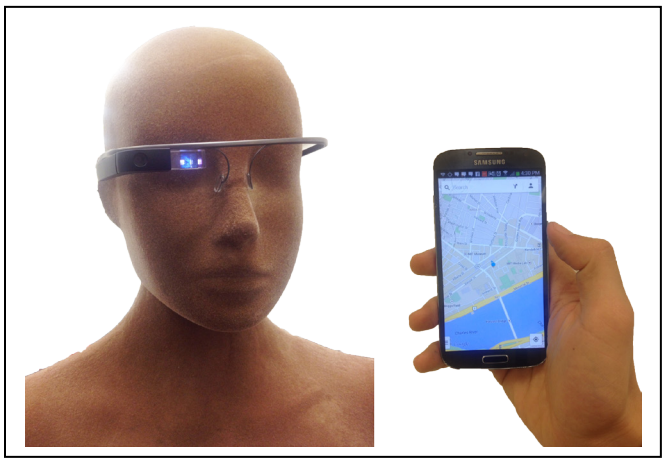

Figure 1. Photographs of the two devices used in this study: Google Glass (left) and the Samsung Galaxy S4 (right).

Procedure

Participants read and signed an informed consent form. Eligibility was verified by interview and a set of questionnaires completed. A brief explanation of the experimental protocol was provided and participants were given detailed introduction on the operation of all three of the navigation entry methods. For each device, participants were first shown how to enter a destination and then given the

Downloaded from pro.sagepub.com by guest on October 18, 2014

opportunity to practice entering an address themselves. Once participants achieved proficiency with all three devices (measured as being able to correctly enter an address), they were trained to drive the simulator and respond to the DRT following the essence of instructions provided in the draft ISO standard. The experimental period was divided into three blocks, one for each destination entry method. The order of the 3 blocks was counterbalanced across participants.

Each destination entry block consisted of training and evaluation segments. During the training segment, participants were first given training on completing the three different tasks (responding to the DRT, driving the simulator, and entering a destination) in an incremental, pair-wise manner. Participants were then trained on completing all three tasks together. Once proficiency on the triple task was achieved (successfully responding to > 70% of the DRT stimuli while also completing a destination entry task and driving the simulator), the evaluation portion of the block began.

In the evaluation segments, participants were provided three minutes of single task driving to establish comfortable control of the vehicle. Following this period, participants were prompted via recorded instructions to enter an address consisting of a 3 digit street number, a street name, a city name, and a state name (e.g. 385 Prospect Street, Cambridge, Massachusetts). After the device finished processing and displayed directions, the destination entry task was considered complete and the DRT was disabled. A buffer period of 30 seconds of driving without other tasks followed, then 60 seconds of driving with the DRT to obtain baseline reaction data (herein referred to as the "just drive" condition), followed by another 30-second buffer. Participants were then prompted with a second, different address. When the second destination entry was completed, participants were instructed to stop the vehicle. Two workload ratings for the destination entry task were then completed: a single global workload rating, 0 (low) – 10 (high), and the NASA TLX scale (Hart 2006). This procedure was repeated for each destination entry device block.

Data Reduction and Analysis

Task duration, standard deviation of lateral lane position and longitudinal velocity, and DRT metrics for each entry method were calculated as the mean across both task trials. The "just drive" period represents the mean value for baseline driving with DRT interval across all three evaluation blocks. Reaction times greater than 1000 milliseconds in this study were counted as a "miss". Miss percentage was calculated as the ratio of number of misses divided by total number of DRT stimuli for each task period. Statistical analyses were performed in R using Friedman's test to assess differences across conditions and the Wilcoxon signed rank test for post-hoc pair-wise comparisons. All error bars and confidence intervals represent the mean-adjusted standard error.

RESULTS

Figure 2 displays subjective workload ratings by destination entry device / method. The global workload ratings are displayed alongside those from the NASA TLX. NASA TLX data was unavailable for two participants. Both the global workload and NASA TLX scales show highly significant effects ($\chi^2(2) = 33.02$, $p < 0.001$ and $\chi^2(2) = 29.55$, $p < 0.001$ respectively). Post-hoc tests reveal significant differences between conditions except between Google Glass and Samsung Voice for both workload scales. A Wilcoxon signed-rank test shows that there is no significant difference between the global workload scale and the NASA TLX scale ($V = 154$, $p = 0.388$).

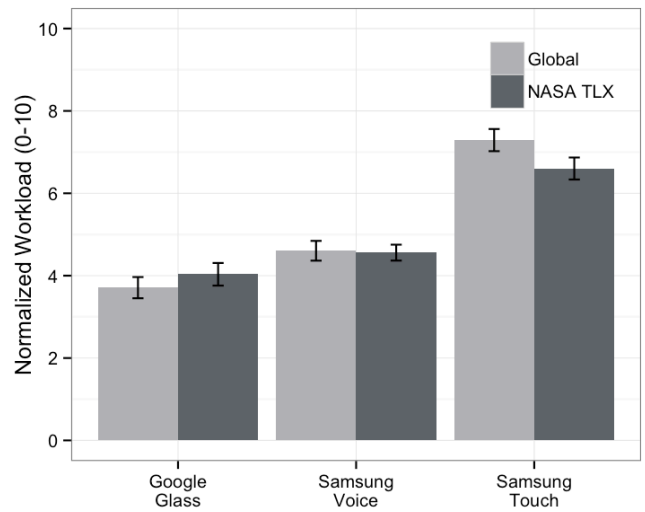

Figure 2. Subjective workload ratings for each destination entry method. The NASA TLX scale, normally ranging from 0-100, is shown normalized for values between 0-10 for comparison to the global workload rating.

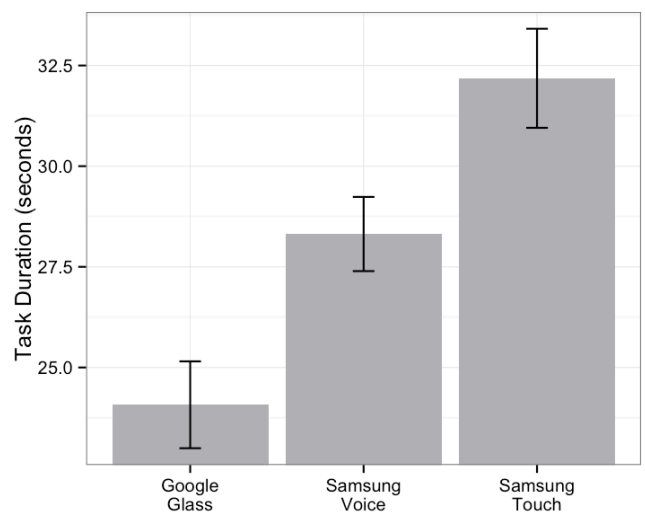

Figure 3. Average task completion time for all three entry methods.

Figure 3 displays task completion time for all three destination entry methods. Task completion was measured as the time from when the driver was prompted to begin the destination entry task to when the device finished calculating directions. Across the three methods, a significant effect is observed ($\chi^2(2) = 18.58$, $p < 0.001$). Furthermore, post-hoc tests show significant differences between Google Glass and both Samsung Voice and Samsung Touch, indicating that destination entry on the smartphone, using both the voice and

Downloaded from pro.sagepub.com by guest on October 18, 2014

touch interface, had a larger time requirement than Google Glass.

Standard deviation of lateral lane position during the destination entry tasks and the "just drive" condition is shown in Figure 4. A highly significant effect is found across conditions ($\chi^2(3) = 30.9$, $p < 0.001$). Furthermore, post-hoc tests show significant differences between all conditions (all $p < 0.001$), except between Google Glass and Samsung Voice.

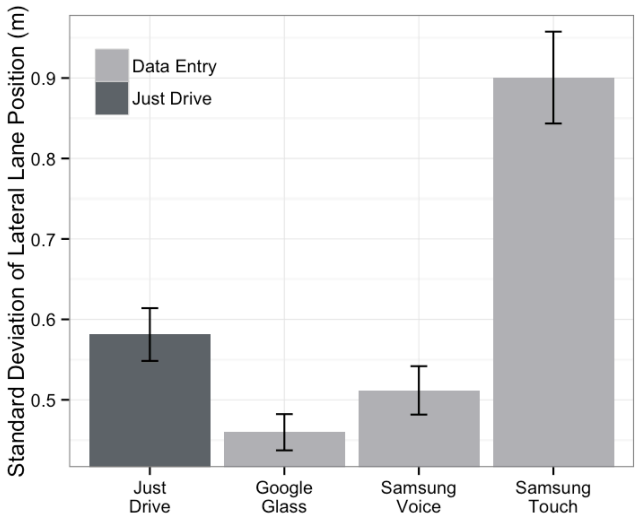

Figure 4. Standard deviation of lateral lane position during destination entry and the baseline driving period ("just drive")

Standard deviation of longitudinal velocity was measured during destination entry for all three devices as well as the baseline driving period. The comparison across all conditions using this metric revealed no statistically significant differences ($\chi^2(3) = 6.2$, $p = 0.1023$). Note that data for the longitudinal velocity was unavailable for one participant.

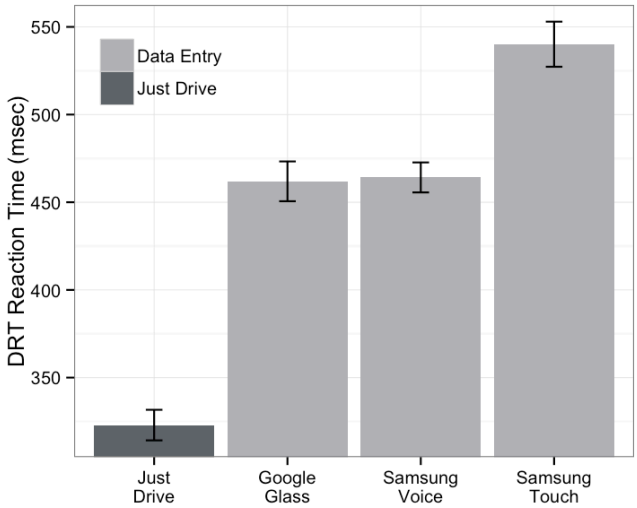

Figure 5. Average reaction time in milliseconds to the red light Detection Response Task (DRT)

Mean DRT reaction time is displayed in Figure 5. The effect across conditions is highly significant ($\chi^2(3) = 50.75$, $p < 0.001$). Post-hoc tests show that Samsung Touch is significantly different from both the Google Glass and Samsung Voice. Paired-comparisons between conditions also show significant differences between the "just drive" period and the three destination entry conditions.

The miss percentage in response to the DRT stimuli is shown in Figure 6. A significant effect is found across the conditions ($\chi^2(3) = 36.69$, $p < 0.001$). Post-hoc tests show significant differences between all three entry methods ($p < 0.025$). Furthermore, the miss percentage during the "just-drive" condition is significantly different from Google Glass and Samsung Touch, but not the Samsung Voice condition.

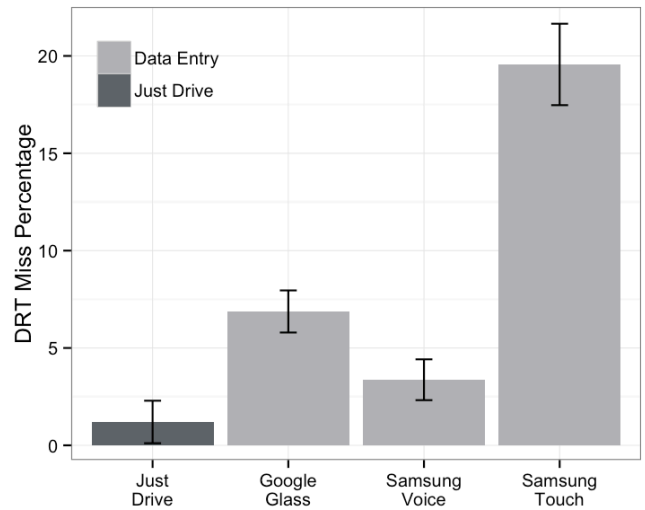

Figure 6. Miss percentage for DRT stimuli

DISCUSSION

This comparison of Google Glass and the Samsung Voice and Touch entry methods revealed some consistent patterns. First, we observed statistically significant differences between the voice interfaces (Google Glass and Samsung) and the Samsung Touch interface. Notably, the voice entry methods differed from the touch interface in being associated with lower subjective workload ratings, apparently better lateral control (lower standard deviation of lateral lane position), shorter task durations, faster DRT reaction times, and lower DRT miss rates.

The differences in subjective workload and DRT reaction time and miss percentage rates indicate that users experienced higher workload while using the touch interface compared to Google Glass. Our findings show that, while many of the participants had years of experience using touch-interface smartphones, after a small amount of training on Google Glass, they found it easier to use than the smartphone touch interface. The subjective workload difference appears consistent with Ranney et al. (2000), who determined that participants experienced lower subjective workload with voice interface devices than touch interface devices. Furthermore, the DRT-based workload data matches the findings of Tijerina, Parmer, and Goodman (1998) in that voice interfaces usually result in a lower workload when compared to their touch interface counterparts. The overall results show lower workload measures for Google Glass when compared to the Samsung touch interface.

The observed increase in miss percentage from 6.9±1.1% with Google Glass to 19.6±2.1% with the Samsung Touch

Downloaded from pro.sagepub.com by guest on October 18, 2014

interface has further implications. In our experiment, the red DRT light was positioned such that it appeared in the driving scene in the relative location of a lead vehicles' brake light. If one were to assume that this was a brake light, results indicate that drivers using a smartphone touch interface would fail to respond approximately 2.8 times as often as those using Google Glass. In comparing these results to other efforts that have utilized the DRT, it is important to note that a "miss" was defined as taking longer than 1000 milliseconds to respond.

A decrease in standard deviation of lateral lane position was observed during the use of the Google Glass as compared to the Samsung Touch interface, but no such trend was apparent in the longitudinal performance metric. Given that Ranney (2008) notes that driving safety is strongly related to lateral and longitudinal control, as well as object/event detection, these combined results suggest that using Google Glass for navigation purposes might not contribute to potential driving risk to the same extent as a traditional touch-interface smartphone.

When comparing the two voice-involved interfaces, the only statistically significant differences between Google Glass and the Samsung Voice interface were in DRT miss percentage ($p = 0.01$) and task completion time ($p < 0.001$). We observed that, across the sample, a participant's miss percentage using Google Glass (6.9±1.1%) was 2 times higher than their miss percentage while using the Samsung Voice interface (3.4±1.0%). These results indicate that participants' per unit time distraction appears somewhat higher using Google Glass for navigation entry than the Samsung Voice interface. Interestingly, the direction of effects for DRT misses and subjective workload, diverge. The subjective workload ratings in Figure 2 show that users thought Google Glass was easier to use, while the DRT miss rate in Figure 6 shows that participants missed the DRT signal more often while using Google Glass. Perhaps, while the Google Glass and the Samsung Voice interface produced similar perceived workload levels, the novel design of Google Glass resulted in a higher level of cognitive absorption.

Returning to the differences in task completion time, this is most likely due to inherent differences in the destination entry methods, such as the number of steps involved with destination entry using each device and method. Mean task duration was: 24.1±1.1 sec (Google Glass), 28.3±1.0 sec (Samsung Voice), and 32.2±1.2 sec (Samsung Touch). This suggests that drivers may be under increased distraction levels but for a shorter period of time when using Google Glass for destination entry vs. using the smartphone voice interface studied; whether one or the other equates to an overall safer driving experience is an open question. The fact that most participants had more familiarity with using a smartphone for destination entry must also be kept in mind.

In conclusion, this research demonstrated that destination entry utilizing Google Glass was less demanding than the Samsung Touch interface as assessed through event detection, driving performance, and subjective workload ratings. Entry through Google Glass did, however, have a statistically higher impact on event detection, but not driving performance, when compared to the Samsung Voice interface. When considering these results, it is important to note that the participant sample was drawn from a population of highly-educated students, and the degree to which these results generalize across the broader population remains to be investigated. Finally, irrespective of device or modality, destination entry was observed to significantly decrease responsiveness to events in the forward scene (as measured by DRT reaction time) compared to baseline driving and thus remains an area of concern.

ACKNOWLEDGEMENTS

Support for this research was provided by the U.S. DOT's Region One New England University Transportation Center. We gratefully acknowledge the contribution of Alea Mehler in managing participant paperwork and proofreading.

REFERENCES

Dopart, C., Häggman, A., Thornberry, C., Mehler, B., Dobres, J. & Reimer, B. (2013). A Driving Simulation Study Examining Destination Entry with iOS 5 Google Maps and a Garmin Portable GPS System. *Proceedings of the 57th Human Factors and Ergonomics Society Annual Meeting*, pp. 1889–1893.

Hart, S.G. (2006). NASA-task load index (NASA-TLX); 20 years later. *Proceedings of the 50th Human Factors and Ergonomics Society Annual Meeting*, pp. 904–908.

ISO 17488 (draft) (2013). Road Vehicles - Transport information and control systems - Detection Response Task: (New Work Item for ISO TC 22 SC13 WG8).

Oracle (2011). "Opportunity Calling: The Future of Mobile Communications - Take Two." http://www.oracle.com/us/corporate/press/522729

Quaresma, M. (2012). "Assessment of visual demand of typical data entry tasks in automotive navigation systems for iPhone." *Work: A Journal of Prevention, Assessment and Rehabilitation*, 41, 6139–6144.

Ranney, T.A. (2008). *Driver distraction: a review of the current state-of-knowledge.* (No. HS-810 787).

Ranney, T.A, Mazzae, E., & Garrott, R. & Goodman, M.J. (2000). NHTSA driver distraction research: Past, present, and future. *Driver Distraction Internet Forum*.

Reimer, B., Mehler, B., Dobres, J. & Coughlin, J.F. (2013). *The effects of a production level "voice-command" interface on driver behavior: reported workload, physiology, visual attention, and driving performance.* MIT AgeLab Technical Report No. 2013-17A. Massachusetts Institute of Technology, Cambridge, MA.

Reuters (2014). "California woman who drove with Google Glass beats traffic ticket." *Reuters - U.S.* http://www.reuters.com/article/2014/01/17/us-usa-googleglass-trial-dismissal-idUSBREA0F1XR20140117.

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 Congress*, *Seoul, Korea, October 12-16*.

Young, K. & Regan, M. (2003). Driver distraction: A review of the literature and recommendations for countermeasure development. *Proceedings of the 2003 Road Safety Research, Policing and Education Conference, Sydney, Australia*, pp. 24–26.

Downloaded from pro.sagepub.com by guest on October 18, 2014