Comparing the Relative Impact of Smartwatch and Smartphone Use While Driving on Workload, Attention, and Driving Performance

Aubrey Samost[1], David Perlman[1], August G. Domel[2], Bryan Reimer[3*], Bruce Mehler[3], Alea Mehler[3],
Jonathan Dobres[3], Thomas McWilliams[3]
[1]MIT Engineering Systems Division
[2]Harvard University – School of Engineering and Applied Sciences
[3]MIT AgeLab and New England University Transportation Center
*Corresponding Author: reimer@mit.edu

A simulator study evaluated the extent to which the use of a smartwatch to initiate phone calls while driving impacts driver workload, attention, and performance, relative to visual-manual (VM) and auditory-vocal (AV) calling methods on a smartphone. Participants completed four calling tasks using each method while driving in a simulator and completing a remote detection response task (R-DRT). Among the 36 participants evaluated, R-DRT miss rates and reaction time were comparable between AV calling on the smartwatch and smartphone, but significantly higher using VM calling on the smartphone. Participants also exhibited more erratic driving behavior (lane position deviation and major steering wheel reversals) with smartphone VM calling compared to both AV calling methods. Finally, participants rated AV calling on both devices as entailing lower workload than VM calling on the smartphone. Overall, few differences emerged for the metrics reported between voice calling on a smartphone versus a smartwatch while driving.

DOI 10.1177/1541931215591347

INTRODUCTION

As the popularity of smartphones has exploded in recent years, so too has interest in companion devices that expand on their capabilities. In 2014, a new genre of personal electronic device emerged that brings many of the smartphone's core functions to its owner's wrist. Smartwatches, long limited to fictional secret agents, now offer smartphone owners easy access to notifications, navigation instructions, and voice input for placing phone calls, text messaging, and internet searches. Though convenience is a key selling point, it may also represent a significant concern for driving safety advocates. Drivers already tempted to interact with their smartphone will now have the ability to do so directly from their wrist.

A growing body of research has investigated the inherent demands of smartphone use while driving (e.g., Basacik, Reed, & Robbins, 2012; Munger et al., 2014; Reimer, Mehler, Reagan, Kidd, & Dobres, 2015; Ranney, Baldwin, Parmer, Martin, & Elizabeth, 2011), but more limited work has been performed to date looking at the driving safety concerns associated with wearable devices. Beckers et al. (2014) found that participants in a driving simulation experiment showed similar response times but a higher miss rate to a detection response task (DRT) while entering an address into a navigation application using Google Glass's voice recognition function compared to voice entry on a smartphone; conversely, task completion time was shorter with Glass. Sawyer, Finomore, Calvo, and Hancock (2014) found that viewing and responding to text messages while driving using Google Glass was moderately less distracting than using a smartphone, but did not entirely eliminate cognitive distractions. Looking more specifically at smartwatches, Giang, Hoekstra-Atwood, and Donmez (2014) reported initial findings on a small sample of drivers' use of a smartwatch to read notifications, finding that drivers were quicker to glance at notifications displayed on the smartwatch and viewed the notification for a longer period of time, compared to reading the same notifications on a smartphone.

Authorities in the United States, Canada, and the United Kingdom have warned drivers over their use of smartwatches, but concede that their laws, as written, do not clearly prohibit smartwatch use while driving (Wiggers, 2014; The Canadian Press, 2014; BBC, 2014). Given the recent availability of smartwatches that not only offer notifications, but also voice and touchscreen-based inputs for key functions, it seems necessary to examine the potential for increased demand posed by the use of a smartwatch as an interactive interface.

In light of the likelihood of increased use of smartwatches during driving, the authors designed a study to investigate the potential for smartwatch use to impact driver workload, attention, and performance. Using a fixed-base driving simulator, the authors compared the use of a smartwatch while driving to initiate phone calls to manual and voice-command based calling methods directly on a smartphone. The authors hypothesized that the use of a voice-command based calling method on a smartwatch would decrease workload and increase attention and performance, relative to both voice-command and manual calling methods on a smartphone due to procedures that are more intuitive and also require lower levels of manual interaction. The following sections outline methods and results of the authors' study to test this hypothesis.

METHODS

Participants

Participants were drawn from the greater Boston area and required to be: experienced drivers (holding a driver's license for more than three years and driving at least once per week), between the ages of 20 and 29 or 55 and 69, and physically and mentally healthy based on self-report. In total, 43 participants were recruited; seven were excluded from the analysis due to experimental interruptions, technical issues, or voluntary withdrawals for factors such as simulator sickness symptoms. The 36 participants considered in the final results are equally balanced by age group and gender. The study was approved through MIT's institutional review board and participants were compensated $60 following incentive instructions detailed in Reimer, D'Ambrosio, Coughlin, Kafrissen, and Biederman (2006).

Equipment

The study made use of the MIT AgeLab's fixed-base, medium-fidelity driving simulator. Built around the full cab of a 2001 Volkswagen New Beetle, it uses an STISIM-based simulation environment projected on a 2.44m by 1.83m screen mounted in front of the vehicle cab. Steering, throttle, and braking inputs are all performed using the original steering wheel and pedals. Previous work has shown strong correspondence between behaviors observed in the simulator and field data (Wang, Mehler, Reimer, Lammers, D'Ambrosio, & Coughlin, 2010; Reimer & Mehler, 2011).

A CogLens unit (http://coglens.com) was employed to support a remote detection-response task (R-DRT). The setup used a red light-emitting diode (LED) mounted on the windshield in the driver's direct line of sight and followed the draft Iso standard (ISO 17488, 2013) with the exception that a foot switch mounted on the vehicle's left foot rest was implemented in place of a finger switch. The foot activated response was used in in place of a finger response since the finger configuration would have interfered with participants' interactions with the smartwatch. During experimental periods involving the R-DRT, the LED was activated at random intervals of between three and five seconds. Previous studies have validated the use of R-DRTs using a foot pedal to detect differences in cognitive task difficulty in miss percentages and reaction times (Bruyas & Dumont, 2013; Angell, Young, Hankey, & Dingus, 2002).

Step	Smartwatch AV	Smartphone AV	Smartphone VM
1	Rotate wrist to wake up watch screen	Press power button to wake up screen	
2		Swipe up from home button to activate Google Now	Select Contacts
3	Speak: "Okay Google"		Tap Search Icon
4	Speak: "Call Mary Sanders"		Begin typing contact's first or last name
5			Select contact's name once visible

Figure 1: Phone call initiation procedures for each calling method studied

The study tested a single model of smartwatch, a Motorola Moto 360, which was paired with a Motorola Droid RAZR M smartphone. The Motorola Moto 360 smartwatch was released in September 2014 and runs Google's Android Wear operating system on a 1.56 inch circular display. The watch is designed to interact with a smartphone via Bluetooth. The Droid RAZR M features a 4.3 inch display and was equipped with Android 4.4, "Kit-Kat." The smartphone was configured to automatically activate the speakerphone option when a call was placed, allowing the researchers to hear when the task had been completed correctly. All calls were routed to a Google Voice mailbox instructing the participant to hang up the call. Participants were instructed to make calls utilizing two calling methods, visual-manual (VM) and auditory-vocal (AV). Figure 1 illustrates the calling procedures for each method included in the study.

Procedures

Participants' eligibility was confirmed with an interview, and they gave informed consent. After consent, they filled-out a pre-experiment survey covering their subjective analysis of their current health and wellbeing as well as their driving habits. The experiment then was broken into three blocks - auditory-vocal calling with the smartwatch (A), visual-manual calling with the smartphone (B), and auditory-vocal calling with the smartphone (C). Block order was randomized and counterbalanced across the gender and age groups. Figure 2 shows the ordering of training and tasks used in the experiment.

Block	Step	Task
A0, B0, C0	1	Informal device training
A-T **B-T** **C-T**	2	Formal training: Calling + DRT
	3	Formal Training: Calling + DRT + Driving *Repeat until successful task completion and >70% response rate on DRT*
A-E **B-E** **C-E**	4	Just Drive without DRT: 60 sec baseline drive
	5	Calling Task: Calling + DRT + Driving
	6	Drive 30 sec
	7	Control: Drive +DRT 60 sec
	8	Drive 30 sec *Repeat for 4 calling tasks and 3 controls*

Figure 2: Experimental procedures for each trial block

Each block was subdivided into training and evaluation portions. Training portions consisted of having the participant place calls while driving, and then increasing the complexity by adding the detection response task. Once a participant could drive, complete the phone call, and respond to 70% of the DRT lights, then they began the evaluation portion. The evaluation portion of each block began with three minutes of driving without other tasks. The last 60 seconds of this interval was used as a baseline driving reference period. Participants then completed four calling tasks of increasing difficulty. Two involved calling a contact by name, and two involved specifying a specific phone for a contact with multiple numbers (e.g. mobile, work). Between tasks, participants were instructed to put the phone down wherever they felt most comfortable (but within easy reach) and just drive for 30 seconds followed by one minute of driving and responding to the DRT light as a control for the block. They again had a 30 second buffer of just driving before placing the next call. At the end of each block, participants rated their workload using a global workload scale (see *Variables*) After all of the blocks were finished, the participants completed post-experiment questionnaires. The entire experiment, including intake, training, and debriefing, lasted approximately 2.5 to three hours per participant.

Variables & Analytic Approach

The research team collected data on and analyzed variables corresponding to three areas of interest: workload, distraction, and driving performance. Subjective workload ratings were obtained after each experimental block. Participants were asked to rate their overall workload while placing calls for the task period they had just completed on a single scale from 0 (low) to 10 (high) using a visual analog format following instructions and a procedure that has been found to produce ratings with high correspondence to global workload ratings obtained with the NASA TLX (see Mehler, Kidd, Reimer, Reagan, Dobres & McCartt, 2015). In order to determine the extent of driver distraction, reaction times and the percentage of missed responses on the R-DRT were considered. Finally, several indicators of driving performance were evaluated: standard deviation of lateral lane position, mean velocity, standard deviation of velocity, and major steering wheel reversals (defined as a change in steering input greater than 3 degrees). Workload ratings were compared across the three methods. Comparisons were made between the average of all three device calling tasks and the average of all driving with DRT control periods. DRT response time and percentage missed were analyzed as the average value of each experimental block compared with the average of all control periods with DRT. Finally, all driving metrics were analyzed as baseline driving without DRT versus just driving with the DRT versus the average of all three calling tasks. Therefore, all results are presented in terms of a single factor with three to five levels, depending on the metric in question.

RESULTS

The following sections present results from the 36 participants considered in the final dataset. Overall, successful task completion rates were 86.4% for smartwatch AV calling, 97.1% for smartphone AV calling, and 92.9% for smartphone VM calling, where success is defined as a participant being able to place the call without assistance. None of the participants experienced a crash event (e.g. collision with another virtual vehicle or leaving the roadway).

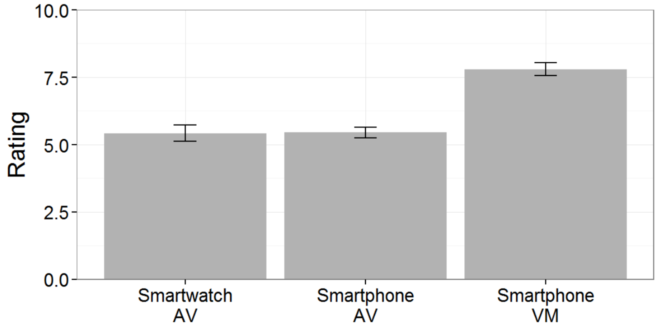

Figure 3: Mean subjective workload ratings from low (0) to high (10) and mean-adjusted standard error bars for the three calling methods.

Figure 3 depicts the mean workload ratings assigned by participants to each method of placing a call. An analysis of variance revealed a significant main effect of method ($F(2,68) = 19.94$, $p < 0.001$), with post-hoc t-tests showing significant difference in workload ratings between smartphone VM and both smartphone and smartwatch AV calling methods ($p < 0.001$).

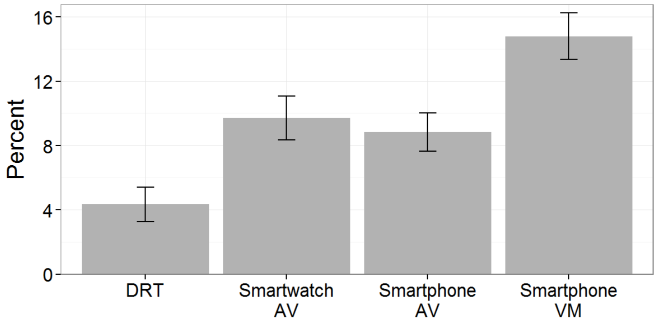

Figure 4: Mean and mean-adjusted standard error of DRT miss percentage by task type

Figure 4 depicts the mean percentage of DRT responses missed by participants along with mean-adjusted standard errors for each calling method and the control period, during which participants were instructed to just drive and respond to the DRT. Analysis of variance showed a significant main effect ($F(3,99) = 8.00$, $p < 0.001$) and post-hoc t-tests revealed that, relative to the baseline periods, smartphone VM ($p < 0.001$), smartphone AV ($p = 0.02$), and smartwatch AV ($p = 0.004$) calling all resulted in higher DRT miss percentages than when driving and not placing a call. They also revealed a higher DRT miss percentage for smartphone VM calling than for AV calling on both the smartphone ($p = 0.006$) and smartwatch ($p = 0.04$).

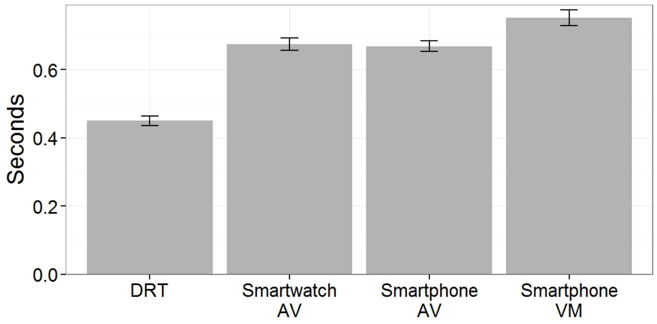

Figure 5: Mean DRT reaction time during calling tasks and mean-adjusted standard error (seconds).

Analysis of variance for DRT reaction time also revealed a significant main effect ($F(3, 99) = 37.30$, $p < 0.001$) and, similar to DRT miss percentage, post-hoc t-tests found that reaction times were shorter during the baseline periods than when initiating a phone call ($p < 0.001$ for all methods). Post-hoc testing also found that smartphone VM calling resulted in longer reaction times than both smartphone AV calling ($p = 0.01$) and smartwatch AV calling ($p = 0.03$). The difference in reaction times between the smartwatch AV and smartphone VM calling methods was not found to be significant ($p = 0.91$).

Finally, in terms of driving performance, analysis of variance did not reveal significant effects with respect to standard deviation of velocity, but did find significant effects with respect to mean velocity (F(4,140) = 3.08, p = 0.02), standard deviation of lateral lane position ($F(4,140) = 11.80$, $p < 0.001$) and major steering wheel reversals ($F(4,140) = 11.98$, $p < 0.001$). In terms of mean velocity, post-hoc t-tests indicate that there was a significant reduction in mean velocity for all task periods relative to baseline (all p<0.02 or greater), but not significant differences in mean velocity between tasks.

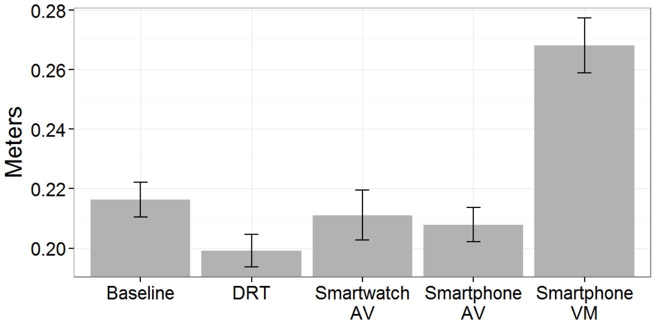

Figure 6: Mean standard deviation of lane position during calling tasks and reference periods (with & without DRT).

Figure 6 depicts the mean standard deviation of lateral lane position during baseline driving without additional tasks (no DRT or calling task), driving with DRT but no calling task, and driving with calling tasks and DRT. Post-hoc t-test results indicate that VM calling on the smartwatch resulted in a significant increase in standard deviation of lane position relative to baseline driving (p<0.001). Values for AV calling on both devices were nominally lower, but not statistically different from baseline. Interestingly, standard deviation of lane position decreased when engaged with the remote DRT but without a calling task, relative to baseline (p<0.01).

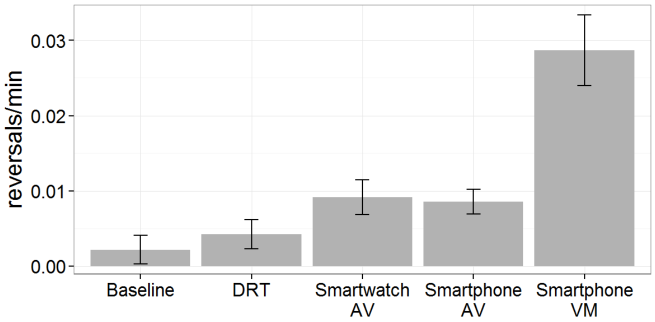

Figure 7: Major steering wheel reversals during calling tasks and reference periods (with and without DRT).

Post-hoc testing found that all three calling methods increased the rate of major steering wheel reversals relative to the baseline period (smartwatch AV p<0.01, smartphone AV p<0.03, smartphone VM $p < 0.001$) (see Figure 7). Further, smartphone VM calling had a higher rate of major steering wheel reversals relative to both AV calling methods ($p < 0.001$ for smartphone AV and $p = 0.005$ for smartwatch AV). Major steering wheel reversals were modestly higher in the DRT only period relative to baseline driving.

DISCUSSION

The results of this study suggest several concerns relating both generally to the use of electronic devices while driving as well as specifically to the use of a smartwatch. Consistent with previous research, participants generally experienced more distraction while initiating interactions with their phones versus just driving. Average DRT miss percentages were almost three times as high during calling tasks than during the control “just drive” periods (11.2 ± 17.7% during calling periods versus 4.0 ± 11.2% during control periods). The difference in DRT reaction time between calling and baseline periods were also greater at 0.70 ± 0.23 seconds compared to 0.45 ± 0.12 seconds, respectively. To the extent that these behaviors are representative of real-world driving situations, these results suggest potential areas of concern. The DRT light placement was meant to serve as a surrogate for relevant stimulus changes in the driving environment, such as brake lights on a leading vehicle, suggesting that drivers engaged in a calling task on their phone, regardless of input method, would fail to notice the brake light of a slowing vehicle ahead twice as often as when they are unoccupied. Moreover, even when they did notice the stimulus light, drivers placing a call on their phone took almost twice as long to respond. Both results suggest the potential for a higher risk of a rear-end collision for drivers who are placing calls.

The observed reduction in standard deviation of lane position during the period when participants were to attend to the windshield mounted remote DRT, but had no other task beyond driving, raises an interesting methodological consideration. This demonstrates a change in a driving performance metric due to the presence of the R-DRT task relative to “just driving”. It is not necessarily surprising that the R-DRT encourages an orientation toward the forward roadway and thus results in a more stable lateral positioning, but it does suggest that it should be kept in mind that engaging in a DRT does fundamentally alter some aspects of the behavior being studied.

In terms of comparisons between the three calling methods, the results clearly indicates that VM calling on the smartphone entails higher workload, is more distracting, and degrades driving performance significantly more than AV calling on both the smartphone and the smartwatch. However, none of the metrics evaluated were found to be significantly different between the two AV methods. Thus, performance with the smartwatch did not appear to differ significantly from smartphone voice calling, except for a nominally higher success rate in placing calls with the smartphone AV interface, suggesting that either represents a preferable alternative to VM calling. That said, none of the participants tested had used a smartwatch prior to participating in the study, but the majority owned a smartphone and were at least familiar with the use of voice calling. Therefore, it is reasonable to suspect that, as smartwatches achieve wider adoption, a participant pool more familiar with the devices might produce different results. It is also worth noting that these results are limited to the Android- and Android Wear-based devices tested, as well as to their application in initiating phone calls. Other devices (e.g., Apple Watch and Apple iPhone) and other tasks (e.g., navigation destination entry or texting) may produce different results.

Conclusion

The results of this study clearly suggest that voice command-based calling methods are preferable to manual calling for the interfaces tested, reinforcing the findings of a number of comparative studies. The between-device comparisons for the AV method did not support the authors' initial hypothesis, as they failed to reveal significant differences in the workload, distraction, or driving performance between smartwatch and smartphone, save for the higher success rate placing calls with the AV smartphone interface. These results do not impact the caution that any diversion of attention from the primary task of driving is potentially risky.

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.

References

Angell, L. S., Young, R. A., Hankey, J. M., & Dingus, T. A. (2002). *An Evaluation of Alternative Methods for Assessing Driver Workload in the Early Development of In-Vehicle Information Systems.* SAE International.

Basacik, D., Reed, N., & Robbins, R. (2012). Smartphone use while driving: a simulator study. (Report No. PPR592) Berkshire, UK: Transport Research Laboratory.

BBC. (2014, September 17). Drivers warned over smartwatch use in cars. Retrieved December 19, 2014, from *BBC News*: http://www.bbc.com/news/technology-29238264

Beckers, N., Schreiner, S., Bertrand, P., Reimer, B., Mehler, B., Munger, D., et al. (2014). Comparing the Demands of Destination Entry using Google Glass and the Samsung Galaxy S4. *Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting*, 2156-2160.

Bruyas, M.-P., & Dumont, L. (2013). Sensitivity of Detection Response Task (DRT) to the Driving Demand and Task Difficulty. *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (pp. 64-70). Iowa City: University of Iowa.

Giang, W. C., Hoekstra-Atwood, L., & Donmez, B. (2014). Driver Engagement in Notifications: A Comparison of Visual-Manual Interaction between Smartwatches and Smartphones. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58*(1), 2161-2165.

International Organization for Standardization. (2013). ISO/DIS 17488 (draft): Road vehicles - Transport information and control systems - Detection-Response Task (DRT) for assessing attentional effects of cognitive load in driving. Geneva, Switzerland: International Organization for Standardization.

Mehler, B., Kidd, D., Reimer, B., Reagan, I., Dobres, J. & McCartt, A. (in press). Multi-modal assessment of on-road demand of voice and manual phone calling and voice navigation entry across two embedded vehicle systems. *Ergonomics*.

Munger, D., Mehler, B., Reimer, B., Dobres, J., Pettinato, A., Pugh, B., et al. (2014). A Simulation Study Examining Smartphone Destination Entry while Driving. *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 1-5). New York: Association for Computing Machinery.

Ranney, T. A., Baldwin, G. S., Parmer, E., Martin, J., & Elizabeth, M. N. (2011). Distraction Effects of Manual Number and Text Entry While Driving. National Highway Traffic Safety Administration. Washington: USDOT.

Reimer, B., & Mehler, B. (2011). The impact of cognitive workload on physiological arousal in young adult drivers: a field study and simulation validation. *Ergonomics, 54*(10), 932-942.

Reimer, B., D'Ambrosio, L. A., Coughlin, J. A., 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., Reagan, I., Kidd, D., & Dobres, J. (2015). Multi-Modal Demands of a Smartphone Used to Place Calls and Enter Addresses during Highway Driving Relative to Two Embedded Systems. Insurance Institute for Highway Safety, Arlington.

Sawyer, B. D., Finomore, V. S., Calvo, A. A., & Hancock, P. (2014). Google Glass: A Driver Distraction Cause or Cure? *Human Factors, 56*(7), 1307-1321.

The Canadian Press. (2014, September 17). Do distracted driving laws cover smartwatches? Retrieved December 19, 2014, from *CBC News*: http://www.cbc.ca/news/canada/toronto/do-distracted-driving-laws-cover-smartwatches-1.2769209

Wang, Y., Mehler, B., Reimer, B., Lammers, V., D'Ambrosio, L. A., & Coughlin, J. F. (2010). The validity of driving simulation for assessing differences between in-vehicle informational interfaces: A comparison with field testing. *Ergonomics, 53*(3), 404-420.

Wiggers, K. (2014, December 16). Incredulous Cop Cites Man for Using His Smartwatch While Driving. Retrieved December 19, 2014, from *Digital Trends*: http://www.digitaltrends.com/mobile/incredulous-cop-cites-man-using-smartwatch-driving/