Texting While Driving: Computerized Eye Tracking as a Method to Assess Visual Distraction During Driving Simulations

Scott Litwiler
Georgetown University

Joseph P. Hasley
Metropolitan State University of Denver

Edgar Maldonado
Metropolitan State University of Denver

Mark Segall
Metropolitan State University of Denver

George Garman
Metropolitan State University of Denver

According to the National Traffic Highway Traffic Safety Administration, distracted driving claimed 3,308 lives in 2022. Few studies have used simulated driving experiences to directly assess the impact on gaze while driving with a cell phone distraction present. In this research, gaze was monitored using a commercial eye tracker mounted to a laptop computer while subjects watched a 10-minute video of normal daytime driving. There was a consistent decrease in objects of interest being observed while text messages appeared. A negative but not statistically significant relationship between the amount of media use (based on responses from the questionnaire) and the amount of time that subjects viewed each text was found (p = 0.072). However, as overall risk for poor driving increased, gaze related to other vehicles decreased (p = 0.014). These results reveal that having a cellphone while driving likely impairs one's ability to notice potentially hazardous conditions on the road that could lead to the harm of the driver or others.

Keywords: distracted driving, eye tracker, driving simulation, road safety

INTRODUCTION

With the increased use of smartphones in today's society it is important to understand the dangers that arise when it comes to operating a motor vehicle while potentially being distracted by a cell phone. According to the National Highway Traffic Safety Administration, distracted driving claimed 3,308 lives in 2021. During that same year, 13% of all traffic fatalities involved young drivers. In one study, for drivers

15–19 years old involved in a fatal car crashes, 21% of the drivers were distracted by the use of cell phones (Khan et al, 2015). A survey of young drivers reported in the same study also found that an estimated 31% of drivers read or sent text or e-mail messages while driving at least once within the prior 30. Previous research has found that the cognitive control hypothesis argues that cognitive load—the amount of stimulus a person is attempting to process at a given moment—impairs driving subtasks (Engstrom, et al., 2017). To build off this research, this study looks more deeply into the specific cognitive task of texting, which is more realistic for individuals driving than results, for example, from Wang's 2010 study that found that horizontal gaze decreased when a subject was presented with a math problem to solve.

This research study was designed to investigate how being distracted by texting affects the salience of various obstacles and conditions experienced by automobile drivers. The goal was to answer the following question: **How does using a cellphone while driving affect the salience of potentially hazardous or informative obstacles?** The prediction was that subjects, when tasked with noticing instances that may be hazardous or informative during driving simulations, will notice fewer instances when presented with distractions from a cellphone. This investigation hypothesizes that increased cognitive tasks decrease a person's gaze distance from the center.

To answer the question, the eye movement of subjects were tracked using a Mirametrix S2 eye tracker while they watched a driving video with sporadic text messages appearing. By comparing the amount of time or speed of noticing relevant instances within the video, the effect of simulated text messaging were estimated. Additionally, information about each subject's driving habits was collected and used for correlative analyses through R.

METHODS

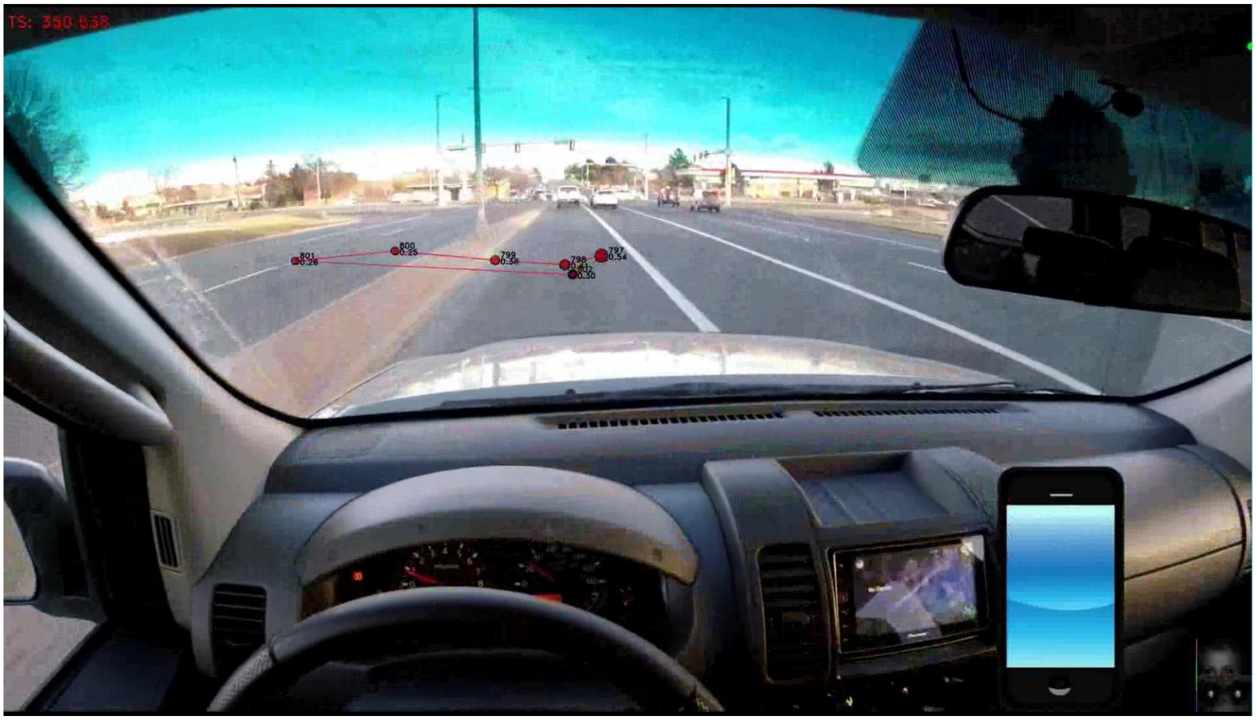
Data were collected from 19 adult subjects. Each subject was assigned a unique numerical SubjectID which was used as an identifier throughout data processing. No uniquely identifying information was retained through data processing. This was to avoid any bias as well as keeping all personal information de-identified.

The driving simulation combined first-person driving video with an overlaid cell phone that would display text messages to the subjects. To record the video a Hero 5 session GoPro was attached to the driver side sun visor of the author's personal vehicle. Video was recorded of the outside of the car, the speedometer, the whole windshield, and the rearview mirror.

Video captures were recorded in multiple environments on data, including residential, urban, and highway settings in the vicinities of Westminster, Broomfield and Boulder Colorado. After all segments were captured, selected segments were edited and combined or discarded. Segments were chosen based on whether the segment had a useable "instance", where an instance was defined as an obstacle or place on the road that could be either informative or dangerous to the driver (e.g., a traffic light, other cars, construction or city signs etc.). Instances were required to fall within a rectangular area within a specific amount of time. This restriction facilitated data analysis for whether a subjects gaze did or did not fall within the region, i.e., they did or did not notice that instance.

The video was edited to consist of 8 segments where instances will occur and each segment will be played twice, once with a phone and once without. When the segment was set to have the cell phone distraction a small blue phone screen would appear in the bottom right corner of the screen. Then every time a text appeared an overlaid sound would ring. All texts were consistent for each subject. Before the driving simulation, a narrative with associated text was shown to each subject. The narrative described what they should do during the video which was to watch the video as if they were driving, determine if the actual driver was a good driver and to answer verbally to any texts that were shown throughout the video. The answers to all text were irrelevant to the research, simply a method of assuring that the subject looked at and comprehended the texts.

FIGURE 1
AN IMAGE OF THE RESULTS RECORDED FROM SUBJECT 1

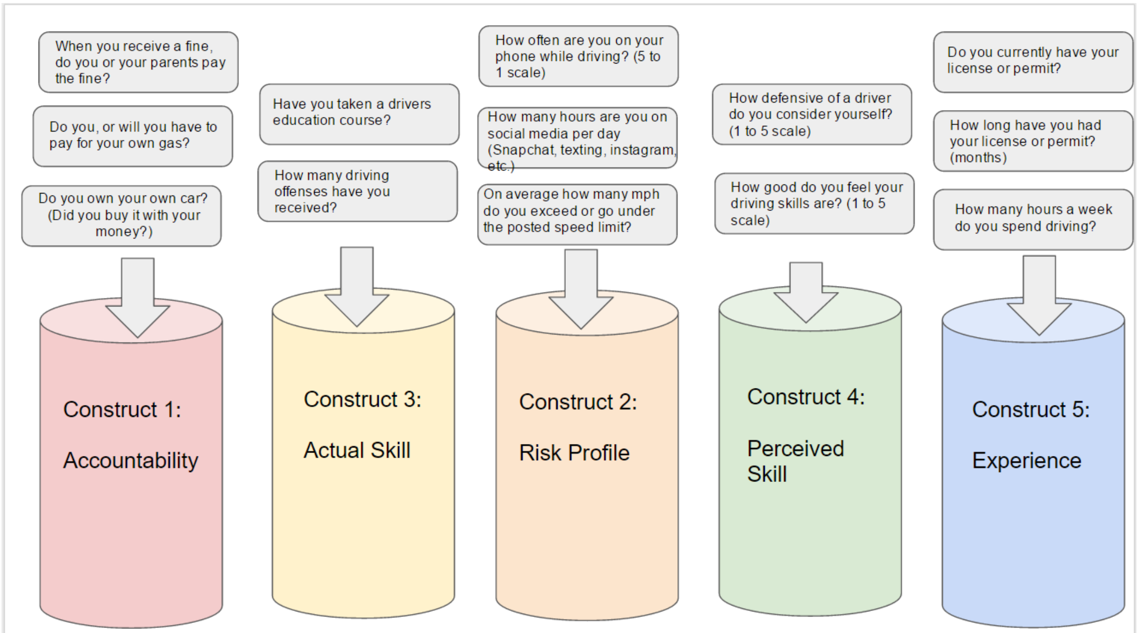


This figure is an image of the video that was shown to all subjects overlayed with the results from subject 1's trial. The red lines and dots on the screen depict where the eyes were looking at that moment in time. This image is a visual for the point of view for subjects, the location of the cell phone, and the video itself.

While analyzing the data, instead of measuring data with respect to time, data was measured with respect to Frames Per Second (FPS). This is done because it made defining regions more accurate. So as opposed to getting measurements to the nearest second it was to the nearest 1/60 of a second because the video was played at 60FPS. The video was always at 60FPS however the eye tracking recording software did not remain constant from subject to subject. This was unchangeable within the software. Therefore the calibration of time for the video was different from each subject's recording. Regions were defined within the video as either texts, cars, or objects and manually set to occur during the video at a specific time. All regions were different sizes based on the instance that was trying to be captured by the region. In addition, when analysing data two variables that were created were PCT and dT. PCT represents the percentage of a subjects point of gaze that fell within a specific region over a given time.

A Google Form questionnaire was completed by each subject after completing the driving simulation. The questionnaire had 17 questions which were designed to probe specific "constructs" related to driving habits or gain demographic information on the driver. Each construct was then tested for correlations with other constructs and with the noticeability of instances. The five constructs were: accountability, perceived skill, actual skill, risk profile, and experience. Each construct was informed by 2-3 questions. The questions that did not fall within a specific construct gave information on the driver such as sex and race.

FIGURE 2
CONSTRUCTS AND THE QUESTIONS THAT FORM THE CONSTRUCT USED TO ANALYZE DATA



Data from the questionnaires were collected using a post-driving test Google form from each subject. The data from the form was exported to a .csv file and then imported into R for processing. For text-based (categorical) data, results were converted to a numeric score. For each question, higher values were assigned for responses that would indicate better or safer driving habits. A subject with a score of 0 would have no accountability, low actual and perceived skill, high risk, and low experience. Results from each question were normalized to a scale of 0 to 1 across subjects and then aggregated into the constructs by adding scores and re-normalized.

All eye tracking data were collected by the S2 Eye Tracker from Mirametrix. The eye tracker was mounted to the top of a Lenovo Ideapad laptop that was used for the driving simulation. All data were collected using the same laptop to keep the resolution and size of the video constant. The sizing of the screen was important because the defined regions of instances were dependent on the specific screen size of the laptop. Efforts were made to minimize distractions during data collection. The data were collected in a quiet room to limit outside distractions and allowed for the subjects to be able to hear each text notification sound. Prior to the start of data collection, the eye tracker was calibrated using Mirametrix software for each subject. If the calibration was not of sufficient quality (i.e., an average error <40), the subject was re-calibrated.

During the driving simulation, x-y data were collected at approximately 60 frames-per-second and saved to the hard drive as a .csv file. There files were the source data for processing.

RESULTS

There were 11 males and 8 females who participated in this study. Data that was collected by the eye tracker included, fixation point of gaze (left and right eye), fixation start and end time, point of gaze (left, right eye and best), pupil diameter (left and right eye), eye position, and screen specifics e.g., dimensions and resolution. For this experiment only the best point of gaze was kept.

While using the eye tracker it was found that the positioning of the subject and the eye tracker played a large role in collecting high quality data. For example, data were best collected when the subject’s head remained still and their eyes were eye level with the tracker. In addition, the tracker was capable of calculating the distance between the tracker and the subject. It was found that calibration was better if data were collected from a subject that was sitting on the far end of the recommended range.

Although the length of the driving video remained constant, the length of the recording for eye tracking for each subject was variable due an inconsistent frame capture rate. This created an inconsistent alignment between regions of interest in time and the recording of the subject's eye movement. To create an alignment eye tracking data and the video were merged using actual time instead of frames.

To ensure all subjects were lined up with the video with respect to start time, they were asked to stare at four balls at the beginning of the video which could then be easily observed as a starting point for each subject that was universal for all subjects. This was then used to calculate a Frame Offset. The Frame Offset is a unique value that is found for every subject that lined up the video with their data collected in time. To calculate the Frame Offset and to ensure a suitable x-y calibration a four-ball method was used. During the calibration four balls appeared in 4 separate quadrants of the screen for three seconds each. Subjects were instructed to stare directly at each of the balls. Calibration data based on the video segment of the four balls was assessed graphically for quality and Frame Offset. If the tracker could not detect which quadrant of the screen the subject was looking in then it was evident that it would not be able to detect whether the subject is looking in a specific region or instance during the course of the video. In addition, by looking at the density plots of subjects' gaze throughout the course of the video, the quality of the data could be determined. Poor data would have a density plot with a shape of a circle with very few points in the region of the cell phone and more centralized or very scattered. As a result of these assessments subjects 4, 16, 17, and 19's eye tracking recording data were excluded.

FIGURE 3A

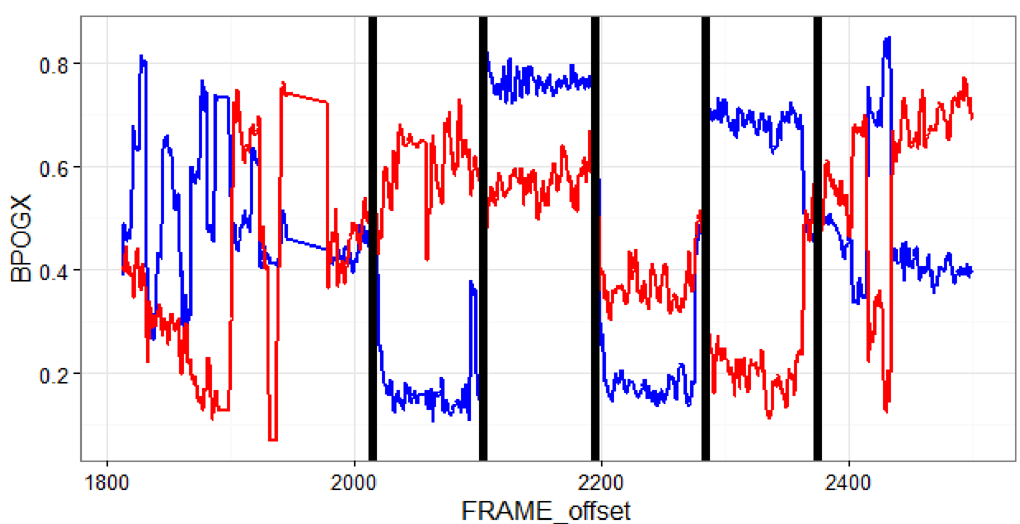
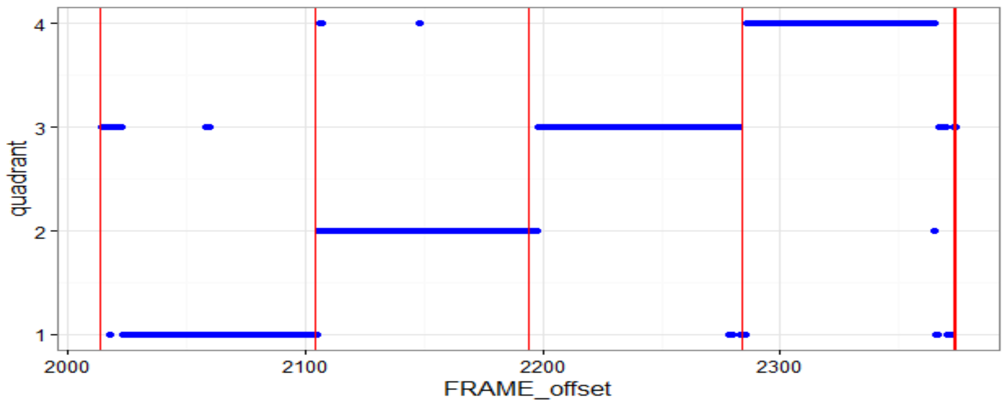


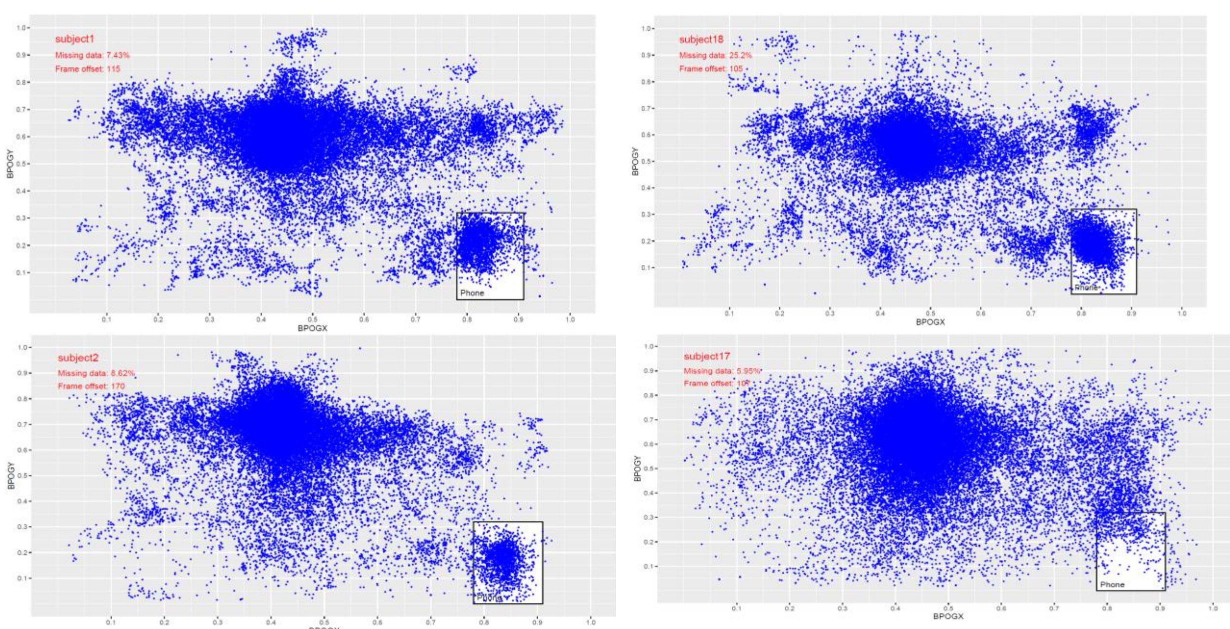
FIGURE 3B



This figure was evaluated when setting a subject's frame offset. In A the blue lines represent the subject's gaze on the X-axis. Red is for the Y-axis. The black lines represent the frame offset and they should fall between each large change in the X or Y directions. Figure B shows which quadrant each observation falls into. For a well calibrated subject, points should sequentially fall into quadrants 1, 2, 3 and 4. With the frame offset set to be between each jump.

The results can be most easily previewed in the density plots . These plots show all points of a subjects' gaze over the duration of the video. The majority of each subjects' gaze fell into the middle of the screen but also spread outwards, showing that the subject was scanning side to side throughout the video. In addition, another place where most subjects gaze fell was in the region of the cell phone meaning they were paying attention to texts. The region that encloses all texts is three seconds long and the mean PCT is 31.1%. Meaning that while the region was turned on for three seconds 31.1% of the time the subjects gaze was in the region. With perfect calibration and data the PCT of the four balls would be 100%, because subjects are told to stare directly at the balls, however the PCT actually came to 56.5%.

FIGURE 4
DENSITY PLOTS OF SUBJECTS 1, 2, 17 AND 18



The plots show the areas most looked at by these subjects. In the lower right corner is the set region for the phone. The first three are examples of good data. The plot in the lower right corner is an example of poor data

The other variable was dT. This represents the time between when a region was turned on and when the first point of a subjects' gaze fell within the region. In this graph positive values at text instances indicate that the subject viewed the phone when a text was present. Negative values at other nontext instances depict that the subject was looking more at the cell phone than the instance that was being recorded. When the PCT values for all text messages for each subject were compared to the PCT values for the calibrations balls, it was clear that the text message viewing time was less than the calibration maximum set by the balls.

All correlations were tested using linear mixed effect models in R. For linear regressions with no covariates, there were 16 degrees of freedom. The effect size was estimated from the slope of the regression. Although most results within constructs were not statistically significantly correlated, there were 2 cases a positive correlation was observed. The most significant correlation was between "actual skill" and "experience". Actual skill was the aggregate of responses related to whether a driver's education course had been taken and the number of offenses received. The experience was related to whether the driver had a license or a permit, how long they had been driving and the number of hours per week that they typically were driving, p=0.00002. This indicated that there was a highly significant correlation between experience and actual skill.

There was also a correlation between "perceived skill" and "actual skill". Perceived skill was based on answers to self-evaluating questions about how defensive and how good a driver each subject thought they were, p=0.005, again highly significant. The subjects who believed that they were good drivers used less media and drove more slowly.

There were 16 text messages throughout the video. The fraction of time over a 3 second period that the gaze was on the defined region for the phone was typically positive. On average, the gaze of subjects was ~55% that of the calibration time of the calibration balls and the average time spent looking at texts was ~30%.

After testing for correlation within constructs, the effect of the phone and instances was tested for correlation with the constructs. First, there was almost a significant correlation between media use and the text time (%)(or the PCT). Here media use is referring to the number of hours an individual spends using social media per day. Using a linear effect model of media use versus text time (%), p=0.072 which is greater than the desired 0.05, however it is very close and with more data a prediction could be made that a correlation would occur. Another prediction that could be made for why this occurred is that as an individual increases their media use they are more used to using a phone and the functionality of a phone, this could cause a decrease in the time they spend looking at the phone.

There was also a correlation between the aggregate of all the constructs and the car viewing time (%). The car viewing time percentage is the same as PCT except instead of being the percentage looked at texts, it is the percentage looked at instances that were cars. The aggregate of all the constructs together creates a 0-1 scale, where 1 would be considered all the good characteristics of a driver determined by their answers to the questions. The second graph on the right depicts that there is a correlation between the aggregate of all constructs and the car viewing time (%), p=0.014. Meaning that if a subject was identified as a good driver, then they would notice less of the potentially hazardous cars on the road. A prediction as to why this correlation occurred is that if someone is considered to be a good driver (having a value of 1 on the aggregated construct scale) they would not need to be as attentive to the road and potentially hazardous obstacles or moments around them to be as safe as subjects defined as bad drivers.

FIGURE 5
DIFFERENCE IN PERCENT OF OBSERVATIONS FOR A DEFINED ZONE BY SUBJECT

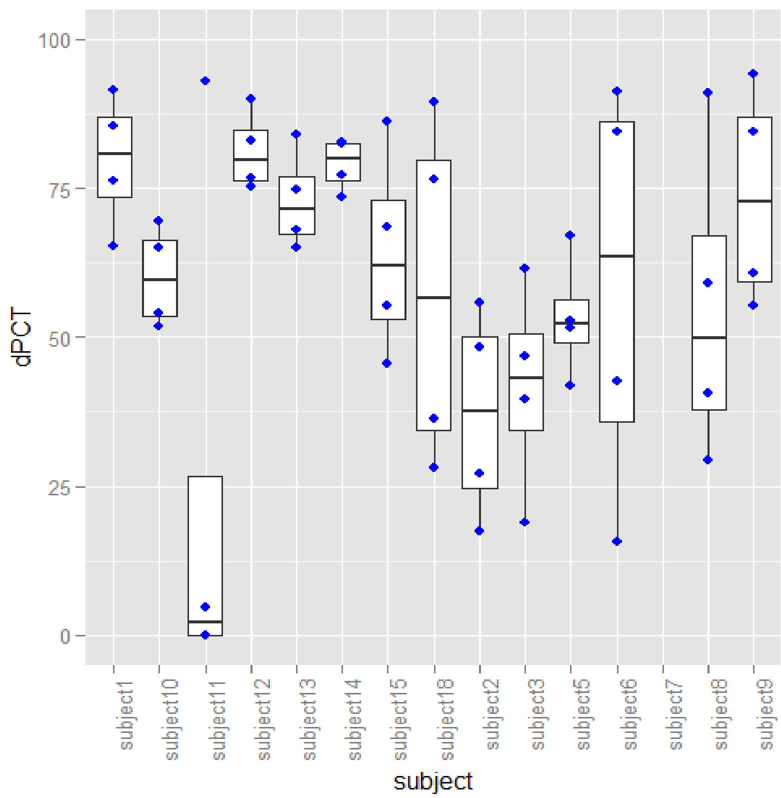
Positive values at text instances indicate that the subject viewed the phone when a text was present. Negative values at other nontext instances depict that the subject was looking more at the cell phone than the instance that was being recorded.

The final correlation found was between a subject's risk profile and the car viewing time(%). Car viewing time is the amount of time that a subject looked in a region that was defined as a car. Correlation was found using a linear effect model. The risk profile of a subject is on a 0-1 scale, 1 being high risk while on the road. This number for each subject is determined by their answers to the questions that fall into the Risk Profile construct, p=0.015. This means that a more risky driver is more likely not to notice or pay attention to the cars that were defined as potentially hazardous within the video.

DISCUSSION

In this preliminary study, real-time eye tracking was used to monitor the effects of simulated text messaging while driving. Data were collected from 19 subjects along with answers to questions relevant to driving habits. Results were studied for correlations between tracking data, constructs related to driving habits and pre-defined zones in the field of vision during the driving simulation. Based on these data it became clear that the calibration and accuracy of the S2 Mirametrix eye tracker was limiting with respect to accurate eye tracking in visual zones of interest during the driving simulation and a more accurate eye tracker would prove to be very beneficial to the research. The use of R is a very reliable way to process and analysis statistical data because of how efficient, effective and flexible it is. Relationships between constructs were found and the appearance of simulated text messages could consistently be detected using the tracking system. There was a reduction in the observation and viewing time of objects of interest (e.g., other cars) decreased when text messages were appearing. Finally, there was a statistically significant correlation between the construct related to high risk driving and the impact of text messaging.

FIGURE 6



The fraction of time (%) that the subject's gaze was within the defined zone of each of the 4 calibration balls at the beginning of the video. This is a visual for the calibration accuracy for each subject.

CONCLUSIONS

This study confirmed that when a person is driving while distracted by a cell phone, they will notice fewer hazardous instances on the road. In addition, it was found that when a driver uses a phone for more time throughout the day they will not spend as long looking at it while driving. Cell phones take the driver's focus away from the road which could lead to harmful and potentially fatal crashes. These data support limiting the use of cellphones while driving and operating a car which supports the common perception of texting and driving, especially in teens who are just starting out.

LIMITATIONS AND FUTURE STUDY

The current literature on texting while driving points to the effects of texting on the number of fatalities, the number of crashes and other measures of poor driving. However, little is reported about the extent of how texting while driving affects a driver’s attention to relevant objects in the driver’s field of vision while driving. The methods for this research study were built from scratch. Much was learned about the accuracy and calibration of the eye tracker and its uses to record information, as well as how to simply process the data. In future studies it would be to the benefit of the research to use a more accurate and sensitive eye tracker to record the data. In addition to adding more subjects to increase the power to detect statistical significance, there should be an increase in the number of questions and changes the wording of the questions. Data should be processed and analysed multiple ways to more effectively detect correlations. Finally, with continuation on this research, the detection regions should be adjusted to maximize gaze detection. The time frame of the regions should also be changed so the PCT data for each region are maximized. If the region is too short in number of frames then it will be 100% excluding error in the tracking sensitivity and if it is too long, the PCT will decrease and show that the subject did not notice the instance when in reality they noticed it but just did not stare at it for the full time that was defined by the region.

REFERENCES

Engstrom, J., Markkula, G., Victor, T., & Merat, N. (2017). Effects of cognitive load on driving performance: The cognitive control hypothesis. *Human Factors*, *59*(5), 734–764.

Kahn, C.A., Cisneros, V., Lotfipour, S., Imani, G., & Chakravarthy, B. (2015). Distracted Driving, A Major Preventable Cause of Motor Vehicle Collisions. “Just Hang Up and Drive.” *Western Journal of Emergency Medicine*, *16*(7), 1033–1036.

National Highway Traffic Safety Administration (NHTSA). (2023). *Traffic Safety Facts: Young Drivers*. Washington, DC. #DOT HS 811 218.

National Highway Traffic Safety Administration (NHTSA). (n.d.). *Distracted Driving*. Retrieved from https://www.nhtsa.gov/risky-driving/distracted-driving.

RStudio Team. (2015). *RStudio: Integrated Development for R. RStudio, Inc., Boston, MA*. Retrieved from http://www.rstudio.com/

Wang, Y. (2010). *The Impact of Repeated Cognitive Tasks on Driving Performance and Visual Attention*. MIT AgeLab & New England University Transportation Center Researchgate.