Examining Driver Distraction as a Function of Driving Speed: An Observational Study using Disruptive Technology and Naturalistic Data

Kentaro Iio*
Zachry Department of Civil & Environmental Engineering
Texas A&M University, College Station, TX 77843-3136
Email: iio_kentaro@tamu.edu

Xiaoyu Guo
Texas A&M Transportation Institute
1111 RELLIS Parkway, Bryan, TX, 77807
Email: xiaoyuguo@tamu.edu

Dominique Lord, Ph.D.
Zachry Department of Civil & Environmental Engineering
Texas A&M University, College Station, TX 77843-3136
Email: d-lord@tamu.edu

Word Count: 4,538 words + 2 tables (250 words per table) = 5,038 words

Submitted: August 1, 2019
ABSTRACT
Considering the number of people who have lost their lives or were injured in crashes associated with driver distractions, it is constructive to understand the prevalent conditions of distracted driving. However, little is known about factors associated with driver distractions because driver distractions are rarely reported until a crash occurs and are likely to be underreported in crash reports. Experiments are limited in terms of validity whereas observational studies are often conducted under limited conditions, such as good weather, daytime, and weekdays. To better understand the characteristics of driver distractions, the authors investigated the relationships between driving speed, posted speed limits, and phone handling frequency through naturalistic driving data (via disruptive technology) obtained on Texas highways by a mobile application (mobile app). As a measure of distractions, a phone handling rate (times/hours driven) was calculated at a 1-mph increment. The phone handling rate was negatively correlated to both driving speed ($r = -0.79$) and posted speed limits ($r = -0.87$). The finding was in line with the theories of risk compensation and boredom as well as other observational studies conducted at fixed locations. Although this research design did not reveal causations, the findings provided new insights into the factors associated with distracted driving. Spatial analyses should be conducted to get a more thorough picture of such factors on crash risk.

Keywords: Distracted Driving, Distraction, Speed, Remote Sensing, Geographic Information System
INTRODUCTION

Distraction is a type of driver inattention that diverts a driver’s focus from the driving tasks to another activity (1). Besides impaired driving and drowsy driving, it is one of the “preventable” crash contributors as indicated a phrase “one life lost is too many” (2). According to the National Highway Traffic Safety Administration (NHTSA), there were about 885,000 estimated distraction-affected crashes in the United States in 2015 (3). The figure accounted for 14.1 percent of all crashes in the year (3). Another report states that crashes involving distracted drivers cost 39.7 billion United States dollars in 2010 in the United States alone (4).

Distractions can take multiple forms, such as visual (e.g., browsing websites), cognitive (e.g., chatting and mind-wandering), manual or physical (e.g., eating and drinking), and auditory (e.g., loud music) (5, 6). In either form, experiments indicate that distractions increase drivers’ mental workload (7), decrease driving performances (8, 9), and are associated with greater crash risks (10, 11, 12). Olson et al. (13) estimated that drivers who use text message while driving were 23 times more likely to face a “safety-critical event” whereas Strayer et al. (14) claim that a cell phone operations while driving can impair driving as much as intoxicated driving does.

In what situations are drivers likely to get distracted? When drivers engage in non-driving tasks during driving, compensatory behaviors called “risk compensation” can take place (15). In this context, theories indicate that drivers are more likely to engage in secondary task when their perceived risk is relatively low (16, 17) or within a certain threshold (18). For example, drivers may reduce their speed when they perceive relatively large risk whereas they may also lower the engagement in non-driving tasks to maintain the perceived safety in driving. As some people have reported that they get distracted when they feel bored during driving (19) or touch their mobile phones (20), there is a chance that the feeling of “boredom” should also be understood in the context of risk compensation (21). Although driving simulators have revealed risk compensations (22, 23), it is not clear if compensatory driving behaviors actually mediate driver distractions in real traffic.

Considering thousands of lives and billions of economic values in the United States alone have been lost due to distraction-affected crashes, understanding the characteristics of distracted driving is an urgent need for today’s society. However, the prevalence of distracted driving is elusive. Although distractions can be reported in crash reports, the number of distracted driving crashes may be underreported because of several reasons (24). First, not all drivers self-report that they were distracted when they are asked. When distracted drivers do not report distractions, it is often difficult or too costly to prove that they were distracted. In addition, different jurisdictions have different crash reporting formats, definitions, and policies, which can affect the likelihood that distractions are properly recorded. In fact, reported distractions as a crash contributing factor has a large variance between jurisdictions. For instance, the Texas Department of Transportation (TxDOT) reports that in 2017, 9.8 percent of crashes in Houston, Texas were classified as “distracted driving crashes” while the rate was 48.5 percent in San Antonio (25). A study using in-vehicle cameras even found that 68.3 percent of drivers were engaged in observable distractions when they crashed (26). Given the situations above, crash reports leave indeterminacy on how often drivers are distracted, regardless of whether distractions are crash contributors or not.

To understand the prevalence of driver distractions, observations can provide partial but more naturalistic information on the prevalence of driver distractions. Although observations may not be able to capture certain types of distractions (e.g., mental and auditory), observers can record visible distractions such as phone handling. In the United States, the NHTSA conducts the
National Occupant Protection Use Survey (NOPUS), which has counted visible phone handling behaviors at randomly selected stop-controlled or signalized intersections annually since 2006 (27). In a 2017 survey, the organization reports that 2.9 percent of drivers were handling cell phones while 2.0 percent of drivers were visibly manipulating handheld devices (28). In the survey, five percent of the drivers were engaged in visible phone handling behaviors behind the wheel.

To understand driving behaviors, other interests lie in about when and where drivers are likely to get distracted. In a systematic review of 51 observational studies on secondary task engagement behind the wheel, Huemer et al. (29) lists driver age, type of road, and vehicle movement as the most influential factors on secondary task engagement rates. In particular, higher secondary task engagement rates were observed among younger drivers than older drivers, in city centers compared to the surrounding regions, and when vehicles were stopped. While observational studies provide information on distraction prevalence, the results can be biased because observations are often made in limited conditions, such as good weather, daytime, and weekdays. Accordingly, so are research syntheses. Although causations were unknown, it is interesting that some observational studies have reported lower distraction rates in urban and suburban areas than rural areas (27, 30).

If conducted carefully, driver distractions could be partially observable inside of vehicles. In fact, some mobile phone application companies have used mobile phones to collect user distraction frequency using built-in gyroscopes. Some reports show the difference in average distraction frequency among drivers in different states (31), metropolitan areas (32), and cities (33). For example, Zendrive (32) reports a higher distraction rate in Houston, Texas (9.44 percent of driving time) and Denver-Aurora, Colorado (8.43 percent of driving time) than in Seattle, Washington (7.13 percent of driving time). Although these attempts may be innovative, the reports lack detailed descriptions in methodologies, descriptive statistics, leaving their research quality in a question. The contexts of distraction, including when, where (on a street level), and why drivers were distracted frequently, also remain not understood. For this reason, it is not clear if the regional differences, if any, provide meaningful insights other than the fact that there might have been regional differences on average distraction rate.

Research Objective and Hypothesis

Taking advantage of the recent development of information technologies, this study intended to analyze the prevalence of driver distractions through naturalistic driving data (disruptive technology) recorded by a mobile application (mobile app). The distraction prevalence was analyzed in relation to driving speed as well as posted speed limits. If drivers adjust their behaviors based on perceived risks, driving speed and posted speed limits are expected to be negatively correlated to a distraction rate. Besides, if there is no difference between speed and phone handling frequency, the following null hypothesis will not be rejected: the driver distraction rate does not change regardless of speed ($H_0$). Other than this hypothesis, this research is exploratory. Because this study does not have an experimental design that involves a control group and an experimental group, the results remain a mere correlation and may not indicate causation. Yet, this research would be valuable because:

1. Little information is available on the relationship between speed and driver distractions.
2. One can see whether or not real driving behavior is in line with the proposed risk compensation theories.
Due to the lack of experimental design, the conclusion of this research would be modest. At best, the authors can say that the results “at least did not contradict” to the theories of risk compensations. Nevertheless, such insights would still be valuable because there is little literature that investigated the topic with real traffic data in an unobtrusive manner. The following section explains the research methodology as well as the characteristics of the datasets.

METHODOLOGY

To investigate the relationships between driving speed and a phone handling rate, the authors analyzed anonymous datasets voluntarily provided by Safe2Save, LLC. The company develops and operates a mobile app that gives users “SAFE 2 SAVE points” for not handling mobile phones behind the wheel (34). Using the global positioning system (GPS), the app records users’ geocoordinates every a few minutes when they are moving. It also records geocoordinates when a motion sensor detects movement larger than a certain threshold when users are moving at 15 mph or faster. This way, the company had the following two datasets recorded by Android devices between 0:00 am on January 1, 2019 and 23:59 pm on June 10, 2019 in Coordinated Universal Time (UTC):

(1) Exposure Dataset – 5,308,632 data points recorded every a few minutes (Mean or M = 4.34 minutes).
(2) Distraction Dataset – 1,125,784 data points created every time users picked up or moved their phones when they were moving at 15 mph or faster but less than 90 mph.

Both datasets had attributes including anonymous user identification number, longitude, latitude, driving speed, and timestamp. In Distraction Dataset, the records were not created when the speed was 90 mph or higher. To align the data range of the two datasets, only data points where speed was 15.5 mph or higher or less than 89.5 mph were used. To eliminate passengers as much as possible, data points of those who were born in 2003 or after were eliminated from the datasets. After this filtering, speed was rounded to the nearest second. Furthermore, posted speed limit data provided by TxDOT (35) were overlaid on these two datasets with a 70-ft buffer by using QGIS 3.8 Zanzibar (36) and PostGIS 2.8.0 (Figure 1). Data points that did not intersect the state-owned roadways were eliminated from further analyses. After this procedure, the Exposure Dataset consisted of 2,445,743 data points whereas the Distraction Dataset had 826,308 data points.
To understand the relative frequency of phone handling events in a given condition, the operational definition of phone handling rate (PHR) was given as Equation 1 and the rate was calculated at a 1-mph interval:

\[
PHR = \frac{\text{the number of data points in Distraction Dataset}}{\text{the number of data points in Exposure Dataset}}
\]  

The larger PHR indicates higher distraction rates. In Equation 1, the denominator represents a time exposure. Since each data point in Exposure Dataset was recorded every a few minutes on average, the number of the data points in a certain area does not accurately represent spatial exposure. Because volume \((v)\) is speed \((s)\) multiplied by density \((d)\), a vehicle moving at 30 mph is twice more likely to be recorded in a certain segment than a vehicle driving at 60 mph would be because the latter vehicle spends a half time of the time the other vehicle spends in that area in a given time.

To examine if there is a difference between the two distributions, a \(\chi^2\) goodness-of-fit test (37) was performed on observed phone handling frequency and expected phone handling frequency based on density functions of observed hours driven by different driving \((n = 74)\). If users are likely to handle their phones regardless of speeds, the two distributions should be identical \((H_0)\). If the null hypothesis is failed to be rejected, that indicates phone handling
frequency did not significantly change by driving speed. If the null hypothesis is rejected, that implies speed or correlated factors, if any, played a role in drivers’ phone handling frequency.

Furthermore, driving speeds in the Exposed Dataset and Distraction Dataset were analyzed in relation to posted speed limits. To investigate the sensitivity of driving speed, 15th percentile, 50th percentile, and 85th percentile speeds were calculated for comparison.

**Methodological Limitations**

The methodology had a couple of limitations. One of them was that the app could not necessarily distinguish passengers from drivers. Although the apps had “passenger unlock” feature that allowed users self-report to stop reporting distractions when they were passengers, there was no guarantee that all passengers self-reported every time they were passengers. The app also could not tell if the users were pedestrians or bicyclists. However, the authors concluded that the noise would have negligible effects as relative aggregated data from the following reasons:

(i) Pedestrians rarely move at 15 mph or faster on state-maintained highways
(ii) In Texas, passenger cars are dominant in relation to buses and bicyclists
(iii) The average occupancy factor for all vehicles in the United States is 1.7 (38), which is less than 2

The second major limitation was the generalizability of the sample. Because SAFE 2 SAVE users might be those who care about safety, the absolute phone handling frequency in this dataset may not represent the general population. However, the analyses within the sample would still reveal phone handling frequency across different speeds when PHR is understood as relative values.

**RESULTS**

Table 1 shows the descriptive statistics of the Exposure Dataset, Distraction Dataset, and PHR. Phone handling events by 11,845 users were observed while 13,360 users left moving footprints.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>n</th>
<th>Mean (mph)</th>
<th>Standard deviation (mph)</th>
<th>Mode (mph)</th>
<th>Min (mph)</th>
<th>Max (mph)</th>
<th>Pace (mph)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure Dataset</td>
<td>2,445,743</td>
<td>57.36</td>
<td>17.01</td>
<td>65</td>
<td>15.50</td>
<td>89.50</td>
<td>64 - 73</td>
</tr>
<tr>
<td>Distraction Dataset</td>
<td>826,308</td>
<td>55.05</td>
<td>18.63</td>
<td>70</td>
<td>15.50</td>
<td>89.50</td>
<td>64 - 73</td>
</tr>
</tbody>
</table>

*Note: Mode and pace were calculated at a 1-mph interval.*

**Driving Speed**

Figure 2 displays histograms of Exposure Dataset (a), Distraction Dataset (b), and PHR (c) by driving speed. While there both (a) and (b) had three small modes at 65 mph, 70 mph, and 75 mph, the two datasets had the same pace (64 mph to 73 mph). Higher PHRs were observed in slower driving speeds. A rejected the null hypothesise between observed phone handling frequency and expected phone handling frequency based on density functions of observed hours
driven at a 99 percent confidence level ($\chi^2 = 77003.33 > 104.01 = \chi^2_{0.001}(73)$). In PHR (c), the mode was found in 16 mph. Driving speed and PHR were negatively correlated ($r = -0.79$).

![Hours driven (n = 2,445,743)](image1)

![Phone handling events (n = 826,308)](image2)

![PHR](image3)

**FIGURE 2** Exposure (a), distractions (b), and PHR (c) by driving speed.

**Posted Speed Limit**

Table 2 shows hours driven, the frequency of phone handling events, and PHR by the posted speed limit. Modes were identified at 55 mph both for hours driven (19,648) and for phone handling event frequency (190,432). The maximum PHR (14.00 times/hour) was observed when the posted speed limit was 35 mph. The authors excluded the 15-mph posted speed limit from further analyses because the observed frequencies were too few. After their elimination, a strong negative correlation ($r = -0.87$) was found between the posted speed limit and PHR, meaning the slower the driving speed was, the more frequent phone handling events were recorded. The negative correlation was stronger between 16 mph and 30 mph ($r = -0.98$) than
between 31 mph and 89 mph ($r = -0.76$). When the authors sampled 80 percent of data from the Exposure Dataset and Distraction Dataset for training and the rest for testing, the PHR was expressed as Equation 2 ($R^2 = 0.98$).

$$\text{PHR} = 0.004x^2 - 0.05x + 23$$

(2)

where: $\text{PHR} =$ estimated PHR

$x =$ driving speed (mph)

Table 2 Hours Driven, Phone Handling Events, and PHR by Posted Speed Limit

<table>
<thead>
<tr>
<th>Posted speed limit (mph)</th>
<th>Hours driven</th>
<th>Phone handling events</th>
<th>PHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>0</td>
<td>1</td>
<td>6.67</td>
</tr>
<tr>
<td>20</td>
<td>3</td>
<td>30</td>
<td>10.75</td>
</tr>
<tr>
<td>25</td>
<td>5</td>
<td>68</td>
<td>13.42</td>
</tr>
<tr>
<td>30</td>
<td>525</td>
<td>7,211</td>
<td>13.72</td>
</tr>
<tr>
<td>35</td>
<td>1,360</td>
<td>19,041</td>
<td>14.00</td>
</tr>
<tr>
<td>40</td>
<td>3,863</td>
<td>51,161</td>
<td>13.25</td>
</tr>
<tr>
<td>45</td>
<td>3,754</td>
<td>43,140</td>
<td>11.49</td>
</tr>
<tr>
<td>50</td>
<td>3,449</td>
<td>35,051</td>
<td>10.16</td>
</tr>
<tr>
<td>55</td>
<td>19,648</td>
<td>190,432</td>
<td>9.69</td>
</tr>
<tr>
<td>60</td>
<td>13,540</td>
<td>106,787</td>
<td>7.89</td>
</tr>
<tr>
<td>65</td>
<td>16,114</td>
<td>112,844</td>
<td>7.00</td>
</tr>
<tr>
<td>70</td>
<td>13,986</td>
<td>106,220</td>
<td>7.59</td>
</tr>
<tr>
<td>75</td>
<td>21,221</td>
<td>149,536</td>
<td>7.05</td>
</tr>
<tr>
<td>80</td>
<td>868</td>
<td>4,710</td>
<td>5.43</td>
</tr>
<tr>
<td>85</td>
<td>10</td>
<td>76</td>
<td>7.75</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>98,346</strong></td>
<td><strong>826,308</strong></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3 plots the 15th percentile, 50th percentile, and 85th percentile recorded speeds in Exposure Dataset and Distraction Dataset by posted speed limits. Although statistical tests were not conducted for posted speed limits, slower mean speed was observed for the 15th percentile ($M = 36.38$, Standard Deviation or $SD = 19.75$) and 50th percentile ($M = 49.66$, $SD = 20.20$) speeds in Distraction Dataset than 15th percentile ($M = 39.11$, $SD = 18.12$) and 50th percentile ($M = 50.85$, $SD = 18.80$) speeds in Exposure Dataset. For the 85th percentile, a drop in mean speed in Distraction Dataset was not observed ($M = 60.29$, $SD = 17.55$ for Exposure Dataset; $M = 60.49$, $SD = 17.33$ for Distraction Dataset). The speed reduction was larger when the posted speed limits were lower than 65 mph ($M = -0.30$, $SD = 1.99$ for 15th percentile; $M = -2.26$, $SD = 4.13$).
0.86 for 50th percentile) than when the posted speed limits were greater than 60 mph ($M = -4.08, SD = 1.53$ for 15th percentile; $M = 0.73, SD = 1.78$ for 50th percentile).

![Figure 3](image.png)

**FIGURE 3** Posted speed limits and recorded driving speed.

**DISCUSSION**

This research explored the relationships between driving speed and phone handling frequency through naturalistic driving data obtained by a mobile app on highways maintained by TxDOT. PHR was defined and calculated as the ratio of the phone handling frequency to time exposure in given driving speed. Although phone handling is not the only form of driver distractions and there was indeterminacy between drivers and passengers, the aggregated data were considered to be valuable because they were obtained unobtrusively.

Because PHR could have included phone handling events by passengers, the absolute value may be too frequent.

**Driving Speed**

When it comes to driving speed (Figure 2), drivers’ time exposure (a) and phone handling events (b) had three peaks at 65 mph, 70 mph, and 75 mph. This might have reflected frequently posted speed limits. The slower the driving speed, the more frequent was the PHR. A $\chi^2$ goodness-of-fit test indicated there was a significant difference between hours driven and the frequency of phone handling events. In addition, a strong negative correlation ($r = -0.79$) was found between driving speed and PHR, meaning that lower driving speeds were associated with higher PHR. The result was not only in line with the existing theories of boredom and risk compensation that have been observed in experimental settings (22, 23), but also compatible with existing observational studies that have reported higher visible distraction behavior engagement rates in urban and suburban areas than in rural areas (27, 30). Obviously, urban and suburban areas tend to have lower speed limits than in rural areas. However, it remains unclear if
higher distractions are mediated mainly by complex visual cues in urban highways or driving speed itself. Therefore, it would be interesting to conduct spatial analyses to identify the dominant factors.

**Posted Speed Limit**

Posted speed limits were strongly and negatively correlated to PHR ($r = -0.87$); namely, lower speed limits were associated with higher PHR. When phone handling events were observed, the 15th percentile and 50th percentile speeds were lower than normal driving conditions (Figure 3). Slower speed profile seemed to be affected. Although this result does not indicate causation, at least it did not contradict to risk compensation theories. It remains unknown as to why the driving speed drop was observed in the Distraction Dataset. Compared to conditions where the posted speed limit was higher than 60 mph, the mean speed reduction was larger when the posted speed limits were 60 mph or less. While the reason remains undetermined, the existence of intersections might have mediated the PHR with the relatively-lower posted speed limits. As well as driving speed, this reason should be analyzed by spatial analyses.

Overall, the findings of this study did not contradict with existing psychological theories. The significance of this research lies with the following two points:

1. It observed vehicle occupants’ relative phone handling frequency by speeds in real traffic.
2. A tendency similar to theories of distractions and boredom was observed: Lower the driving speed and lower the posted limit, the more frequently app users tended to handle their phones. The tendency was more obvious when the driving speed was less than 30 mph and posted speed limits were under 65 mph.

Limitations of this research were noted in the Methodology section above. Because both hours driven and phone handling rates showed unimodal distribution over driving speed, the noise by other modes of travelers seemed to be limited. Even though this paper revealed relationships between driving speed and the prevalence of phone handling events, the reasons behind the results remain unknown due to the lack of experimental design. If one can identify spatial characteristics of distracted driving, this could help extend our knowledge on the issue. Therefore, further investigation should be done through spatial analyses.

**ACKNOWLEDGMENT**
The authors would like to thank Safe2Save, LLC for providing valuable datasets under their terms and conditions.

**AUTHOR CONTRIBUTIONS**
The authors confirm contribution to the paper as follows: study conception and design: Iio, K.; data collection: Iio, K.; analysis and interpretation of results: Iio, K., Guo, X., Lord, D.; draft manuscript preparation: Iio, K., Guo, X., Lord, D. All authors reviewed the results and approved the final version of the manuscript.
REFERENCES


Iio, Guo, and Lord


37. Pearson, Karl. “On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling.” The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 50, 302 (1900): 157-175. doi: 10.1080/14786440009463897