U.S. Department
of Transportation
**National Highway
Traffic Safety
Administration**

NHTSA
www.nhtsa.gov

DOT HS 811 820

August 2013

Text Reading and Text Input Assessment in Support of the NHTSA Visual-Manual Driver Distraction Guidelines

DISCLAIMER

Suggested APA Format Citation:

Boyle, L. N., Lee, J. D., Peng, Y., Ghazizadeh, M., Wu, Y., Miller, E. & Jenness, J. (2013, August). *Text reading and text input assessment in support of the NHTSA visual-manual driver distraction guidelines* (Report No. DOT HS 811 820). Washington, DC: National Highway Traffic Safety Administration.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT HS 811 820		

4. Title and Subtitle
Text Reading and Text Input Assessment in support of the NHTSA Visual-Manual Driver Distraction Guidelines

5. Report Date
August 2013

6. Performing Organization

7. Author(s)
Boyle, L. N., Lee, J. D., Peng, Y., Ghazizadeh, M., Wu, Y., Miller, E… & Jenness, J.

8. Performing Organization

9. Performing Organization Name and Address
Westat 1600 Research Blvd, Rockville, MD, 20850

10. Work Unit No. (TRAIS)

11. Contract or Grant No.
DTNH22-11-D-00237

12. Sponsoring Agency Name and Address
National Highway Traffic Safety Administration
Office of Program Development and Delivery
1200 New Jersey Avenue, SE.
Washington, DC 20590

13. Type of Report and Period
Final Report

14. Sponsoring Agency Code

15. Supplementary Notes
Julie Kang was the Contracting Officer's Technical Representative (COTR) for this project.

16. Abstract
The aim of this project was to provide data supporting the development of NHTSA's proposed Visual-Manual Driver Distraction Guidelines' text entry and text reading specification. The purpose of the study was to examine the two test protocols recommended in the proposed NHTSA Guidelines: the driving simulator and occlusion goggle protocols, under different conditions of text type, text length, and ambient text conditions. In the driving simulator, the total eyes off road time (TSOT) measures the number of seconds drivers eyes left the road to complete the text entry and text reading tasks. Similarly, the total shutter open time (TSOT) measures the total time the participants could view the display through the occlusion goggles. The results of the study indicated that the use of a TSOT of 12 seconds was more appropriate than 9 seconds. In addition, the mean ratio between TSOT and TEORT for the Text Entry task, across all text length and ambient text conditions, was 1.03. The corresponding mean ratio for the Text Reading task was 1.09. NHTSA concluded that these ratios suggest there is no need for a field or inflation factor when comparing TSOT values to TEORT. In summary, the driving simulator and occlusion goggle protocols produce consistent indicators regarding the distraction potential of text entry and text reading tasks.

17. Key Words	18. Distribution Statement
Driver Distraction, Visual Manual Guidelines, HMI, traffic crashes, texting, mobile phone,	Document is available to the public from the National Technical Information Service www.ntis.gov

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22
Unclassified		73	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

ACKNOWLEDGEMENTS

Several undergraduate students helped screen participants, reduce data, and synthesize literature who should be acknowledged in this report. They are Charles Meissner, Zixiang Xuan, Miralis Torres, and Madeleine Gibson.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS, ACRONYMS, AND SYMBOLS

Alliance	Alliance of Automobile Manufacturers
ANOVA	analysis of variance
CMS	changeable message signs
DOT	Department of Transportation
DFD	dynamic following and detection
EOR	eyes-off-road
EOT	enhanced occlusion task
JAMA	Japan Automobile Manufacturers Association
MGD	mean glance duration
NADS	National Advanced Driving Simulator
NHTSA	National Highway Traffic Safety Administration
SD	standard deviation
SMS	Short Message Service
TEORT	total eyes-off-road time
TSOT	total shutter open time
UW	University of Washington
UW-Madison	University of Wisconsin-Madison

EXECUTIVE SUMMARY

The National Highway Traffic Safety Administration is concerned with the effects of distraction due to drivers' use of electronic in-vehicle devices on motor vehicle safety. Many studies show that driver inattention and distraction significantly increase crashes (Horrey & Wickens, 2006; Klauer, Dingus, Neale, Sudweeks, & Ramsay, 2006; Regan, Lee, & Young, 2008). NHTSA has proposed voluntary Driver Distraction Guidelines for Original Equipment Electronics Devices With Visual-Manual Interfaces[1] to encourage the design of interfaces that will minimize distraction when performing secondary tasks while driving.

The aim of this project was to provide data supporting the development of NHTSA's proposed Visual-Manual Driver Distraction Guidelines' text entry and text reading specification. The purpose of the study was to examine the two test protocols recommended in the proposed NHTSA Guidelines: the driving simulator and occlusion goggle protocols, under different conditions of text type (text entry, text reading), text length (three levels: short, medium, long), and ambient text (i.e., text surrounding the text to be read. two levels: present, not present).

The tasks involved participants entering words or reading phrases using a touchscreen while driving in the simulator (UW) or wearing occlusion goggles (UW-Madison). An identical 7'-inch touchscreen display with QWERTY keyboard was used for both experiments. There were 28 participants from four age groups (18-24, 25-39, 40-54, and 55-75 years old) with seven participants in each group. The data collection was conducted using the NADS MiniSim driving simulator and a FaceLab eye tracker located at UW, and CogLens occlusion goggles located at the UW-Madison.

A mixed factorial, complete block design was used for each study with two between-subject (age group, gender), and three within-subject independent variables: task type (two levels: text entry, text reading), text length (three levels: short, medium, long), and ambient text (i.e., text surrounding the text to be read. two levels: present, not present). For text entry, the short, medium, and long text length were 4, 6, and 12 characters, respectively; for text reading, they were 20 to 40, 60 to 80, and 120 to 140 characters, respectively. The proposed NHTSA Guidelines state one repetition for each test condition in the simulator trials, and ISO 16673 indicate five repetitions for each test condition in the occlusion trials. For this study, there were three replications for each test condition in the simulator, occlusion, and static trials. The average experiment time for each participant was two hours.

The dependent variables examined were driver's eyes-off-road (EOR) durations for the driving simulator study and shutter open time for the occlusion study. These included mean duration of individual eye glances away from forward road view (mean glance duration, MGD), the sum of individual eye glance durations away from forward road view (total eyes-off-road time, TEORT),

---

[1] Draft guidelines are available online on Fed. Reg. 77 FR 11199 [Feb. 24, 2012].
, but as of this writing, the final version has not yet been published.

and total shutter open time (TSOT). The proposed NHTSA Guidelines provides four of the five criteria used for the driving simulator and occlusion studies. The fifth criterion is based on discussions with NHTSA (Ranney, Baldwin, Parmer, Martin, & Mazzae, 2012).

Driving Simulator Study

Criterion 1: For at least 21 of the 24 test participants, no more than 15 percent (rounded up) of the total number of eye glances away from the forward road scene have durations of greater than 2.0 seconds while performing the testable task one time.

Criterion 2: For at least 21 of the 24 test participants, the mean duration of all eye glances away from the forward road scene is less than or equal to 2.0 seconds while performing the testable task one time.

Criterion 3: For at least 21 of the 24 test participants, the sum of the durations of each individual participant's eye glances away from the forward road scene is less than or equal to 12.0 seconds while performing the testable task one time.

Occlusion Study

Criterion 4: For at least 21 out of 24 participants, the TSOT should be less than or equal to 9.0 seconds.

Criterion 5: For at least 21 out of 24 participants, TSOT should be less than 12.0 seconds.

The proposed NHTSA Distraction Guidelines state that 85 percent of the participants should conform to the above acceptance criteria. The original acceptance criteria outlined in the guidelines were based on a sample size of 24; however, the current study used 28 participants. Therefore, the number of participants who conform to criteria 2, 3, 4, and 5 was rounded to 24 (85% of 28) to adjust for the extra data collected.

Repeated measures ANOVA was used to assess differences in the dependent variables for all three independent variables (task type, text length, ambient text) and respective interaction terms. The project team also examined the variation in performance across participants to determine what proportion of drivers were able to conform with the criteria specified by NHTSA for the occlusion and driving simulator protocols. In addition to using the guideline criteria, standard deviation and 95 percent confidence intervals were also included when the means were reported.

The results of the simulator study showed significant effects for text length and ambient text on EOR. The MGD and TEORT were significantly longer for the text entry tasks than for text reading tasks with comparable text length tested in the study. Longer text strings resulted in longer MGD and TEORT for both text entry and reading tasks. The ambient text significantly increased TEORT for text reading but the effect size was not as large as text length.

The findings of the occlusion study showed significant differences in TSOT for task type and text length, but no differences were observed for the ambient text condition. Specifically, TSOT was longer for text entry tasks compared to text reading tasks. Consistent with the simulator study, TSOT was significantly longer for the long text entry task when compared to the short and medium text entry task.

The results show that the occlusion and simulator studies generally led to similar outcomes. None of the text entry tasks conformed to acceptance criteria 1 and 2, regardless of length and presence of ambient text. The short text entry tasks conformed to criteria 3, 4, and 5. The medium text entry tasks conformed to criterion 3, and the medium text entry tasks without ambient text conformed to criterion 5. All text reading tasks conformed to criteria 2, 3 and 5, regardless of length and presence of ambient text. Short and medium text reading tasks conformed to criteria 1 and 4.

There were, however, different results observed between the simulator and occlusion study in the ambient text condition. In the simulator study, the TEORT was longer in duration with ambient text than without, but in the occlusion study no TSOT differences were observed. The differences in TEORT between the two protocols may be due in part to an artifact of the protocols themselves. The simulator protocol may result in longer TEORT values because of varying glances needed to maintain road position in the simulator, which is not observed in the occlusion protocol (where no driving was required).

A direct comparison of the secondary tasks between the simulator and occlusion protocol showed that the use of a 12-second criterion for the occlusion TSOT (criterion 5) provided task acceptable results that were more consistent with the 12-second TEORT criterion (criterion 3) for driving glances) than for the 9-second TSOT (Criterion 4). Furthermore, this research showed that the mean ratio between TSOT and TEORT for the text entry task, across all text length and ambient text conditions, was 1.04. The corresponding mean ratio for the text reading task was 1.09.

1. INTRODUCTION

1.1 BACKGROUND

The National Highway Traffic Safety Administration is concerned with the effects of distraction due to drivers' use of electronic in-vehicle devices on motor vehicle safety. Sixteen percent of fatal crashes and 21 percent of injury crashes in 2008 were attributed to driver distraction (Ascone, Lindsey, & Varghese, 2009). The results of the 100-Car Naturalistic Study showed that driver inattention and distraction, including secondary tasks engagement, driving-related inattention to the forward roadway, non-specific eye glances, and fatigue, were associated with 78 percent of crashes and 65 percent of near-crashes, significantly increasing crash risk (Klauer et al., 2006).

NHTSA has proposed voluntary Driver Distraction Guidelines for original equipment electronic devices with visual-manual interfaces[2] (Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices, 2012) to discourage implementation of tasks performed such devices unless the device interfaces are designed to minimize distraction when tasks are performed while driving. The primary objective of the proposed NHTSA Distraction Guidelines is to reduce the number of motor vehicle crashes and the resulting deaths and injuries that occur due to distraction from the primary driving task while performing non-driving activities with the in-vehicle electronic devices. The Guidelines are presented as an aid to manufacturers in designing and evaluating in-vehicle devices that can be safely used and to minimize unsafe behavior resulting from use of the devices.

Driver distraction is defined as the diversion of drivers' attention away from activities critical for safe driving (Lee, Young, & Regan, 2008). There are three primary types of distractions: visual, manual, and cognitive, which can also be described as instances in which drivers take their eyes, hands, or mind off the road. Typically, most distractions involve a combination of visual, manual, and cognitive distraction. Distracting tasks with visual-manual component (all distractions have some cognitive component) contribute to higher crash risk (Klauer et al., 2006; Angell et al., 2006).

Visual-manual distraction, as defined by NHTSA, requires a driver to look at a device, manipulate a device-related control, and/or watch for visual feedback from the device. For the past two decades, NHTSA has studied the effects of distraction on driving safety. More recently, NHTSA has shifted its focus to develop methods and metrics for measuring distraction resulting from in-vehicle electronic device use. This has led to NHTSA proposing voluntary guidelines for minimizing the distraction potential of in-vehicle and portable devices. Although other guidelines exist, such as the Alliance guidelines (Alliance of Automobile Manufacturers, 2006),

[2] Ibid.

NHTSA issued its own guidelines. Among the reasons listed for this decision in the notice proposing the Guidelines were the following:

- Include the latest research on driver distraction, which considers technology unavailable at the time the other guidelines were developed.
- Clarify those aspects (i.e., protocols) that are unclear in the other guidelines.

Consistent with NHTSA's initial focus, this section summarizes some of the more relevant studies conducted on visual-manual distraction to clarify the independent variables (e.g., text type, character lengths), data collection protocols, dependent variables, and performance thresholds (eyes-off-road time) used in this study. This section synthesizes publications that specifically relate to the following research questions:

- How do text entry and text reading distractions affect driver performance?
- How can this information be applied to developing in-vehicle electronic devices?

1.1.1 Text Entry

Ranney et al. (2012) conducted a study that evaluated several text entry related secondary tasks using the Alliance Principle 2.1B evaluation protocol developed by the Alliance of Automobile Manufacturers (Alliance). This protocol focused on driving performance degradation resulting from five secondary tasks (Alliance of Automobile Manufacturers, 2006): dialing a contact, destination entry, 10-digit dialing, radio tuning, and text entry (with pairwise comparisons done for all combinations). Pairwise comparisons show that there were significant differences in the number of lane exceedance (crossing) for all comparisons except dialing a contact versus 10-digit-dialing, and destination entry versus10-digit dialing. Similarly, significant differences in standard deviation of headway were observed in all comparisons except dialing contact versus 10-digit-dialing (Ranney et al., 2012). In general, the findings show that text messaging resulted in the greatest performance degradation (i.e., more lane exceedance and larger standard deviation of headway), followed by destination entry, whereas radio tuning resulted in the least degradation. The other two tasks (10-digit phone dialing and dialing from a contact list) were relatively equal with moderate performance degradation.

Ranney et al. (2012) also noted that there was a strong correlation between task duration and the driving performance metric of lane exceedance. As expected, the five secondary tasks required different completion time. When lane exceedance was adjusted to reflect exceedance per second rather than exceedance per task (Ranney, Baldwin, Parmer, Martin, & Mazzae, 2011b), the differences in the resulting metrics were much smaller and generally insignificant. In fact, only text messaging had significantly larger number of lane exceedance per second compared to dialing a contact, 10-digit-dialing, destination entry, and radio tuning. The latter four tasks were not significantly different from each other once the lane exceedance metric was adjusted by task duration. Ranney et al. (2011b) examined the same set of secondary tasks using the Dynamic Following and Detection (DFD) metrics (i.e., dynamic car following task combines with a visual target detection task), which resulted in slightly different results from Ranney et al. (2012). This study showed that text messaging was most distracting as it resulted in significantly larger SDLP and longer response time to task detection, followed by 10-digit dialing. Radio tuning and destination entry were least distracting. Although destination entry and radio tuning did not show significant differences in driving performance (e.g., standard deviation of lane position and detection task response time), destination entry did take considerably longer task completion time than radio tuning and phone tasks. The longer task duration can expose drivers to additional risk. Performance was similar for touchscreen and hard button phones; however, touchscreen phones were slightly more distracting (Ranney et al., 2011b).

With regard to the comparison of the two evaluation alternatives (Alliance and DFD), the following conclusions were drawn:

- The two implementations of Alliance Principle 2.1B verification procedure produced inconsistent results, despite both implementations being largely consistent with specifications in the Alliance Guidelines. Differences in driving (primary) task demands appear to have contributed to this finding.

- The results imply that protocol specifications in the Alliance Guidelines are not sufficiently detailed to ensure consistent implementation of the verification procedure (Ranney et al., 2012).

The Ranney et al. (2011b, 2012) studies only examined driving performance degradation caused by text entry but did not examine the associated eye glance behavior. The present study will examine drivers' eye glance behaviors as influenced by different types of text entry and reading tasks.

1.1.2 Text Reading

The process of reading has been studied for over 100 years (Rayner, 1998), and these prior studies lay the foundation for understanding the influence that text features have on driving while reading. Much of the previous research has focused on unconstrained reading of text, text presented with a moving window, or text presented tachistoscopically. The effects of spatial and temporal variations on reading marquee text, as might occur while driving, have also been examined. Studies show line length and scroll rate significantly impact reading (Chen & Tsoi, 1988; Duchnicky & Kolers, 1983; Juola, Tiritoglu, & Pleunis, 1995; Kang & Muter, 1989). Findings indicate that these variables affect reading rate (or speed) more than reading comprehension, suggesting that participants may adjust their reading speed to achieve a desired level of comprehension. However, these studies examined reading on a desktop computer as a single task, providing little insight into how reading would be affected when driving. Goals in reading as well as the mental model and context of reading have also been shown to affect reading speed and comprehension (Fletcher & Chrysler, 1990).

Relatively few studies have considered the constrained reading situations that occur when drivers must share their attention between the road and text. Reading text that is unrelated to goals currently instantiated in working memory requires a greater attentional shift than text related to the task at hand. For example, text about relevant road conditions such as messages on road signs may be read more quickly than text unrelated to the driving task such as a song title on a radio display.

Wierwille's (1993) model of visual sampling can describe the primary difference between a reading-only task and reading while driving. As a secondary task, reading from a display is performed in between short glances (1-1.5 s) to the roadway. Figure 1 shows patterns of how awareness changes with driver glances. Long off-road glances may be particularly detrimental because drivers are more likely to be delayed in responding to, or even missing, critical events. The longer an off-road glance, the less aware the driver is of the dynamic driving context. This figure reinforces NHTSA's concern about the consequence of long glances and the need to understand how characteristics of text reading and entry contribute to such long glances.

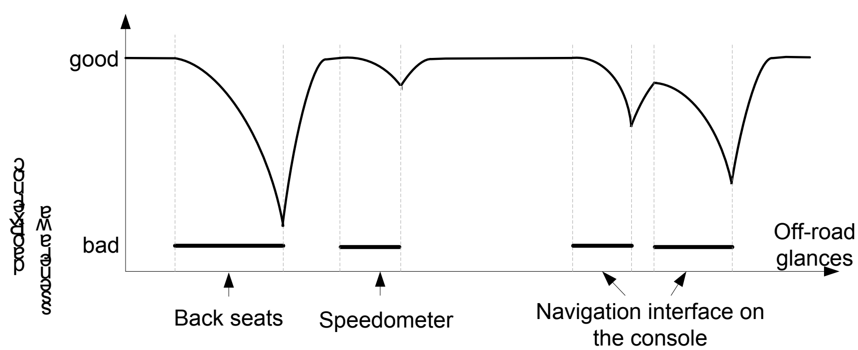


Figure 1. A hypothetical profile of driver's awareness of the road context during the alternation of on- and off-road glances (Liang, Lee, & Yekhshatyan, 2012).

The Alliance Guidelines include several criteria to limit glance durations. Principle 2.1 of the Alliance Guidelines states, "Systems with visual displays should be designed such that the driver can complete the desired task with sequential glances that are brief enough not to adversely affect driving." The criteria for this principle states that single glance durations should not exceed 2 seconds, and total glance time to complete the task should not exceed 20 seconds. The 2-second criterion was determined using Rockwell's (1988) data; a 2-second maximum glance time included the 85^{th} percentile for glances. The 20-second criterion was determined using Dingus et al. (1989) where the 85th percentile included about 10 glances for task completion; 10 glances multiplied by the 2-second maximum equates to a 20-second total glance time (JAMA, 2004).

1.1.3 Factors that Influence Text Entry and Reading

1.1.3.1 Practices in Changeable Message Signs

Proper character rule adjustment is a vital component to convey appropriate messages, while maintaining a desirable quantity of text. This notion is well studied for changeable message signs, also referred to as variable message signs or dynamic message signs. CMSs also have the advantage of displaying messages in context, while most in-vehicle messages can be context free and can range from vehicle status updates (e.g., next exit 30 miles) to fast food advertisements (e.g., McDonalds turn right). A study conducted by the New Jersey Department of Transportation evaluated drivers' understanding of messages using various character and word substitutions. The study concluded that not explicitly specifying days and times led to unacceptable ambiguities, while simple character adjustment like replacing "thru" with a dash was acceptable (Dudek, 1999).

A study conducted by Westat, Inc., on real-time travel information evaluated driver's ability to read and process information presented on CMS. This study used reading time, ease of assimilating information, willingness to change routes, and confidence in route choice as the dependent variables. Findings related to text reading derived from this study (Lerner, Singer, Robinson, Huey, & Jenness, 2009) include:

- Participants took longer to read the sign when there was a timestamp (indicating time of last update of the CMS);
- Drivers took longer to process the sign that showed speed instead of travel time; and
- Drivers took longer to process the sign with color-coded text.

Dudek and his colleagues have conducted extensive research on CMS that is highly applicable to in-vehicle displays (Dudek, 1992, 1997, 1999; Dudek, Schrock, & Ullman, 2007; Dudek, Schrock, Ullman, & Chrysler, 2006). For example, Dudek (1992) discussed scrolling roadside messages on highway changeable message signs and showed that drivers were able to process messages up to 8 words in length without a decrement in comprehension, provided each word was presented with a one-second minimum viewing time. The effect of CMS on driving performance was not discussed in this report.

Practices for CMS messages are further applicable to this study because they are designed to convey the message while minimizing the time a driver must divert their eyes from the roadway, generally designed to present the message in about 8 seconds or less (Dudek, 2004). Research published in the CMS Operation and Messaging handbook lists six key strategies in creating messages that enhance motorist understanding, which are as follows (Dudek, 2004):

- Simplicity of words,
- Brevity,
- Standardized order of words,

- Standardized order of message informational units,
- Widely understood abbreviations when abbreviations are needed, and
- Standardized applications of messages.

The CMS handbook also states that, "an efficient, brief, and to-the-point message is a good message. Just because there are spaces available on a CMS does not mean that all spaces should be used for a message (pg. 2-4)" (Dudek, 2004). This is a critical concept and applicable to in-vehicle electronic devices. Information credibility is also a significant factor and another principal that can be adapted from CMS research. It is important to avoid displaying information that is irrelevant, obvious, repetitive, trivial, and erroneous (Dudek, 2004). In addition to considering character length, units of information or informational unit, is another way to quantify text. Essentially, "one unit of information refers to the answer to a question a motorist might ask. A unit of information is typically one to three words, but at times can be up to four words" (Dudek, 2008). Figure 2 provides an example of this model.

333		
Question	*Answer*	*Unit of Information*
1. What happened?	ACCIDENT	1 unit
2. Where?	PAST ROWLAND	1 unit
3. Who is advisory for?	FAIR PARK	1 unit
4. What is advised?	USE FITZHUGH	1 unit

Figure 2. Units of information in text reading, (Adopted from Dudek, 2008).

1.1.3.2 Chunking

A "chunk is a collection of elements having strong associations with one another, but weak associations with elements within other chunks" (Gobet et al., 2001). This is best understood through the example of a telephone number. By grouping numbers into chunks of information, one can better remember a string of numbers even though additional characters such as dashes and parentheses are included. Many studies have evaluated various components to chunking and how it relates to human comprehension. One study found that "the chunk structure of a telephone number (e.g., for a 7-digit telephone number, this might follow a 3-4 representational structure) provided natural break points to return attention to driving" (Brumby, Salvucci, Mankowski, & Howes, 2007). By utilizing this notion, character classifications and groupings may be able to better address more practical applications for the guidelines.

1.1.3.3 Information bits and number of characters

The JAMA Guidelines offer several recommendations regarding the amount of information to present on a display. JAMA (2004) states that, "the displayed visual information shall be sufficiently small in *volume* to enable the driver to comprehend it in a short time or be presented in portions for the driver to scan them in two or more steps." For the guidelines to be usable by automobile manufacturers, this "volume" should be quantitatively measured to identify acceptable or unacceptable amounts of information.

The JAMA guidelines state that for dynamic information displayed in letters, the number of letters (e.g., Japanese characters [kana - Japanese phonogram], alphabets, numbers) displayed at a time shall not exceed 31 (JAMA, 2004). In the JAMA guidelines, a number (e.g., 120) or a unit (e.g., "km/h") is treated as a single letter, and does not appear to consider the different amount of information provided by Japanese characters (kana), English alphabets, and numbers (JAMA, 2004). However, there are differences in how characters and text is used in different countries, and their respective requirements for human perception. This is especially important to consider when a single kana character can translate to multiple alpha characters in the English alphabet. In fact, there has been research investigating the informational content between kanji (Chinese character), kana (Japanese phonogram), and the English alphabet (Kamiya, Nakamura, & Matsumoto, 1994). According to Kamiya et al (1994), letters and numbers contain approximately 6 bits of information whereas objects/pictures contain 14 bits of information. Hence, the relationship between kana and English alphabet can be compared with respect to bits or amount of information presented. For example, 154 bits of information can be contained in 13 kana letters, one number, and five objects (1 banana, 2 apples, 1 orange, 1 circle) and this is computed as:

$$(13 \text{ letters} + 1 \text{ number}) \times 6 \text{ bits} + 5 \text{ objects} \times 14 \text{ bits} = 154 \text{ bits}$$

Several studies have shown that one English alphabet letter corresponds to approximately 2.2 bits (Shannon, 1951; Yavuz, 1974). Additionally, a kanji character can represent 10.8 bits and a kana can represent 6 bits (Kamiya et al., 1994). Using these values, an evaluation was performed to map 30 Japanese characters with two different kanji to kana ratios to the equivalent number of English letters. It was observed that 30 Japanese characters, containing 24 kanji letters (80%) and 6 kana letters (20%), could be equivalent to 134 English alphabet characters. If the 30 Japanese characters contain 12 kanji and 18 kana letters (kanji to kana ratio = 40%: 60%), then this would correspond to 108 English alphabet characters. Therefore, 30 Japanese characters with kanji to kana ratios ranging from 40%:60% to 80%:20% could correspond to 108 to 134 English alphabet characters.

1.1.3.4 Ambient Text

Ambient text is defined for this project as text that is not relevant to the targeted text reading or text entry task but exists within the in-vehicle display in an unspecified location. A workshop held at the NHTSA Vehicle Research and Test Center on March 23, 2012, discussed the concerns related to the 30-character limit and "where should the 30-character limit be applied to in a display." There are many ways that ambient text can be seen on an in-vehicle display, and even text located as time and date, or on touchscreen control dials that can be considered ambient text. However, these latter bits of information tend to be consistently located in the same position. When relevant information appears more consistently in a given information channel or area of interest, observers will tend to sample this channel more often, i.e., look at particular areas when they expect to find relevant information (Horrey, Wickens, & Consalus, 2006). Therefore, the hypothesis for the current study is that the unspecified location of ambient text (Figure 3) is more likely to draw a driver's attention than ambient text that is consistently displayed at a known location. The literature also recommends the use of clear, simple fonts and reduction of ambient text to only what is essential (Campbell et al., 1998).

Figure 3. Examples of in-vehicle displays.

1.1.3.5 Two-Second Eye Glance Duration Threshold of Distraction

The test method proposed in the NHTSA Distraction Guidelines limits distraction by restricting the amount of time that a driver's eyes are away from the roadway while performing non-driving related tasks. One of the key measures for this assessment is examining eye glances away from the forward view. Off-road glances associated with secondary tasks divert the driver's visual attention from the driving situation and impose visual distractions. The initial analysis of the 100-car study showed that the sum of off-road glance durations exceeding 2 seconds in 5 seconds before and 1 second after the onset of the precipitating event increased the risk of crashes/near-crashes by approximately two times (Klauer et al., 2006). Controlled experiments also showed that long off-road glances (e.g., >2 seconds) lead to larger lane deviation and slower response to lead vehicle braking (Dingus et al., 1989; Donmez, Boyle, & Lee, 2007b). Glances to an in-vehicle display that is located further away from the center of the road can lead to a slower response to hazardous events (Horrey & Wickens, 2004; Lamble, Laakso, & Summala, 1999).

Rockwell (1988) conducted three studies over a 6-year period with 106 individuals, consisting of 200 highway drives of 45 to 60 minutes each. In this study, subjects performed various tasks using different in-vehicle stereo designs, in addition to sampling various commanded and natural glances to side mirrors. Over 6,000 off-road glances for stereo tasks were recorded, as well as

4,000 natural and commanded speedometers and mirror glances. Rockwell (1988) observed that the number of glances was a much more sensitive measure than average glance duration for examining driver-specific characteristics and display-control design. Errors made by the driver usually did not affect the average glance duration (which ranged from 1.27 to 1.42 sec) but did lead to an increased number of glances. The author observed that, "*when complex displays require glance durations beyond the 90th percentile, most drivers are clearly facing special visual workload problems. If information cannot be resolved in the average glance duration due to legibility problems, the driver may be tempted to increase [average glance duration] to perform the task, thus compromising his disposition, safety or both"* (Rockwell, 1988, p. 322). Therefore, a safe task should be completed using glances under the 90th percentile, which corresponds to about 2 seconds. The major conclusion of the paper, as stated by the author, is that the glance duration "*is impacted more by the demands of the driving task than by 'in car' targets and their visual characteristics*." This was supported by a notable 20-percent decrease in glance duration when there were more demanding road situations (increase traffic, more difficult road attributes). Moreover, Rockwell (1988) concluded that "*poor display/control design is usually reflected in more glances, not longer glances*."

Wierwille, Antin, Dingus, and Hulse (1988) conducted a study with 32 drivers who were asked to complete 26 tasks (e.g., remaining fuel, defrost, and navigation system tasks) of varying difficulty. The analysis focused on total glance time. Most notable, they found that mean glance durations for completing any task, regardless of perceived difficulty, ranged from 0.62 to 1.63 seconds. However, the number of glances varied greatly from 1.31 glances to 6.91 glances. Though no statistical analysis was conducted on single glance durations or number of glances, the findings were consistent with Rockwell (1988) in that secondary tasks had more of an impact on number of glances than glance duration. Moreover, the data showed that the driver's mean glance duration was typically below two seconds, regardless of the number of glances.

1.1.4 Occlusion Study Protocol

One approach for evaluating the visual demands of in-vehicle devices is to use a surrogate task. Occlusion goggles provide a means to control for the intermittent view of drivers when they are driving and performing a secondary visual-manual task. The ISO standard 16673 specifies the number of rules for assessing visual distraction using an occlusion procedure. According to the ISO standard 16673, at least 10 participants should be used in evaluating each system configuration. In addition to the typical experimentation protocol (e.g., use licensed drivers, provide instructions on use of apparatus, practice task), the vision-occlusion intervals enforced by the goggles should be set at 1.5-second shutter open and 1.5-second shutter closed cycles (ISO, 2007). However, both Alliance and JAMA guidelines state using cycle times of 1.5-second open/1.0-second closed (Alliance of Automobile Manufacturers, 2006; JAMA, 2004). In regard to secondary task acceptance criteria, the Alliance indicates that the TSOT must be less than 15s and JAMA says TSOT must be less than 7.5s. The aim of the occlusion method here is to identify designs where text reading and text entry tasks can be accomplished in a few, short glances.

Ranney et al. (2011a) conducted three experiments for evaluating driver distraction. Two of these experiments are directly related to the forthcoming work. One experiment compared the traditional occlusion protocol with an enhanced occlusion protocol. The traditional protocol asks a participant to complete a secondary task while wearing occlusion goggles. The three tasks included destination entry by address, destination entry by selecting a previous destination, and finding a designated city within a long list generated by the navigation system. At specific intervals, the occlusion goggles would obscure the participant's view of the secondary task, thereby simulating the need to attend to the road (i.e., one cannot focus on both the secondary task and the road). That said, the participants could continue working on the secondary task even when the occlusion goggles completely obscured their view.

A second experiment used an enhanced occlusion task (EOT) that incorporated an auditory tracking task during occluded periods. The auditory tracking task required participants to move a joystick in response to tones presented in one of two audio channels (left or right), thereby providing the same type of processing load required to steer a vehicle, but without the visual demands that could conflict with the mechanisms of the occlusion protocol. The experiment asked participants to complete three different tasks in three different conditions (static – no occlusion, occlusion, EOT) using a within-subject design. It was hypothesized that the EOT would increase the amount of time required by the participant to complete a task by impacting the "resumability metric," or ability for the participant to resume the secondary task once completing the auditory task (Ranney et al., 2011b). Although there was an increase in the time participants took to complete a task using the EOT when compared to the traditional occlusion protocol, some blind operation of the secondary task (as described earlier) still occurred (11%). Moreover, the completion time required for secondary task in both traditional occlusion and EOT was less than the total time required by the "static" trial. Hence, although EOT was designed to minimize blind completion of tasks, it was not perfectly achieved. The problem is complicated because the durations of complex tasks differ considerably (Ranney et al., 2011b). The protocol of the third experiment was not used in our current study, but it did examine the distraction effects of three navigation systems in which the driver completed two secondary tasks: (1) selecting a previous destination by scrolling from a list of address and (2) destination entry by typing an address.

There are two general methods for occluding driver's vision. The more common method is having the participants wear occlusion goggles that have lenses composed of liquid crystal displays that can rapidly switch between transparent and opaque, based on a signal from a computer. An example of a device using this method is PLATO Visual Occlusion Spectacles by Translucent Technologies. The other method involves mounting a shutter on top of the task display to occlude the display, based on commands from a computer. There are no differences between two methods in terms of task timing and driving performance. A review show several studies that use an occlusion technique select the timing of occlusion intervals to mimic the glance times in actual driving situations and the need to interrupt the task. The glance times ranged from 1 to 2 seconds and the occlusion times ranged from 1 to 5 seconds in most studies (Tsimhoni, 2003).

Harbluk, Burns, Go, & Morton, (2006) used PLATO goggles (by Translucent Technologies) to assess the distraction potential of a number of navigation tasked performed using four production vehicle navigation systems. In accordance with the 2005 draft version of the ISO 16673 standard, the shutter open time to shutter close time ratio in this study was set at 1.5:1.0 (in seconds) (ISO, 2005). Participants completed two types of navigation tasks: address destination entry, and point of interest entry. Two levels of complexity were considered for each task, i.e., low complexity, where a previously stored destination was selected, and high complexity, where a new destination was entered. The experiments were all conducted in a stationary vehicle. The results showed that for all four navigation systems tested, high complexity tasks did not conform to the 15-second maximum time criterion for total viewing time. Low complexity tasks, on the other hand, conformed with the acceptable viewing time criterion (Harbluk, Burns, Go, & Morton, 2006).

1.1.5 Summary of Background Review

Several guidelines to provide recommendations and best practice of in-vehicle display design have been available prior the publication of the proposed NHTSA Driver Distraction Guidelines. However, they might not all be suitable for all driving situations in the United States. The JAMA Guidelines provide very narrow constraints for implementation in vehicles used in Japan that may not be applicable in the United States (e.g., the 30-character rule). There are also differences in recommendations (e.g., Alliance, JAMA, SAE) for examining the factors related to distraction associated with text reading and text entry. Although previous studies have examined the effect of different types of text entry and reading tasks on driving performance, few studies have examined the same task with variable lengths specifically to assess design limitations. Eyes-off-road and occlusion time have been used as measures of distraction and can provide quantitative comparisons among different design specifications, and thus inform appropriate design.

1.2 OBJECTIVE AND RESEARCH QUESTIONS

The goal of this project was to provide data supporting the proposed NHTSA Visual-Manual Driver Distraction Guidelines' text entry and text reading specification. The purpose of the study was to examine the two test protocols recommended in the proposed NHTSA Guidelines: the driving simulator and occlusion goggles, under different conditions of text type (text entry, text reading), text length (three levels: short, medium, long), and ambient text (i.e., text surrounding the text to be read. two levels: present, not present).

The project team evaluated the values selected for text reading and text input, as described in the proposed NHTSA Distraction Guidelines associated with non-driving tasks. The values set forth in the guidelines are based on work from several publications that are available from SAE, ISO International Standards, and on the NHTSA Web site (www.nhtsa.gov/Research/Human+Factors/Distraction).

The project team used a driving simulator and the occlusion technique. In the two study protocols, text that drivers may typically read and enter while driving was used with in-vehicle displays. The tasks included entering text such as for destination entry in a navigation system and reading text typically observed on changeable message signs. The dependent measure for the driving simulator study was EOR time and for the occlusion study was TSOT.

Specifically, the proposed NHTSA Distraction Guidelines (Section X. Option EGDS and Option OCC) defines acceptance criteria for the simulator and occlusion testing:

Eye Glance Testing Using a Driving Simulator:

> Criterion 1: For at least 21 of the 24 test participants, no more than 15 percent (rounded up) of the total number of eye glances away from the forward road scene have durations of greater than 2.0 seconds while performing the testable task one time.
>
> Criterion 2: For at least 21 of the 24 test participants, the mean duration of all eye glances away from the forward road scene is less than or equal to 2.0 seconds while performing the testable task one time.
>
> Criterion 3: For at least 21 of the 24 test participants, the sum of the durations of each individual participant's eye glances away from the forward road scene is less than or equal to 12.0 seconds while performing the testable task one time.

Occlusion Testing:

> Criterion 4: For at least 21 out of the total 24 participants (85% rounded to a whole number), the TSOT should be less than 9.0 seconds.
>
> Criterion 5: For at least 21 out of the total 24 participants (85% rounded to a whole number), theTSOT should be less than 12 seconds. This criterion provides a more direct comparison to the EOR time criterion used in the simulator test.

2. METHODS

2.1 SAMPLING AND PARTICIPANT RECRUITMENT

This study used 28 participants with valid driver's license recruited from Seattle for the driving simulator study and another 28 participants recruited from Madison, Wisconsin, for the occlusion study. In Washington, six participants were replaced due to data loss (a total of 34 participants were recruited to obtain complete data for 28 participants). Similarly, in Wisconsin four participants were replaced to compensate for data loss (a total of 32 participants were recruited to obtain data for 28 participants). The use of 28 participants is a deviation from the 24 participants outlined in the proposed NHTSA Distraction Guidelines. The project team had received additional funds to run one more participant in each age group per study site in order to increase the sample size.

The inclusion and exclusion criteria were determined based on previous experiences on simulator studies and the proposed NHTSA Distraction Guidelines (Section X. VI.4). Each participant had to meet the following criteria to be included in the study:

- Be in good general health (no heart condition, seizure, epilepsy, Ménière's disease, or narcolepsy);
- Be an active driver with a valid State-issued driver's license;
- Drive a minimum of 7,000 miles per year;
- Be in the age range of 18 to 75 ;
- Be comfortable using computer and touchscreens;
- Be comfortable communicating via text messages;
- No participation in any driver simulator studies in the past 6 months; and
- Be a native English speaker.

In addition, participants who used any special equipment to drive (i.e., booster seats, pedal extensions, hand brake or throttle, spinner wheel knobs, or seat cushions) or identify themselves as having a high likelihood of experiencing simulator sickness were excluded from participating the simulator study. Participants who have poor eye calibration results from the eye tracker were also excluded in order to reduce the loss of data due to unusable eye glance data. For the occlusion study, the flickering of the occlusion goggles might increase the chance of those prone to epileptic seizures. Although this is a very rare occurrence, those who are prone to seizures were excluded.

For both studies, the study participants were recruited via campus e-mails, flyers, and online advertisements. Interested individuals who contacted the research team were first screened via telephone. Individuals who were willing to participate and met all inclusion criteria were placed into the study. Participants were compensated $20 per hour for their participation. For the driving simulator study, participants who drove to the UW campus were also provided with parking validation for their visits. Approximately 100 people were screened for each study to meet the intended number of participants.

The participants were representative of the U.S. driving population and were drawn from four age groups (18-24, 25-39, 40-54, and 55-75 years old) per the proposed NHTSA Distraction Guidelines. For the driving simulator study, seven participants were used in each age group (Table 1).

Table 1. Participant Gender and Age by Age Group (UW, Driving simulator study)

Age groups	**18-24**	**25-39**	**40-54**	**55-75**
Mean Age (yrs)	21.1 (0.4)	28.7 (4.3)	48.6 (3.1)	59.7(3.0)
Gender				
Male	4	3	4	4
Female	3	4	3	3

Similar to the driving simulator study, seven participants from each of the same four age groups participated in the occlusion study (Table 2).

Table 2. Participant Gender and Age by Age Group (UW-Madison, Occlusion study)

Age groups	**18-24**	**25-39**	**40-54**	**55-75**
Mean Age (yrs)	21.3 (0.8)	30.4 (4.0)	46.6 (3.6)	61.9 (5.6)
Gender				
Male	3	4	3	4
Female	4	3	4	3

2.2 EXPERIMENTAL PROTOCOL

2.2.1 Experiment Facilities

Data collection for the driving simulator study at UW was conducted using the NADS MiniSim, a low-cost, low fidelity driving simulator, and Seeing Machines faceLAB eye tracker located at UW. The NADS MiniSim includes three screens (3.0 ft. wide by 1.7ft. tall, each) that are placed about 4.5 ft. away from driver's eye point. The faceLAB eye tracker has the capability to characterize test participants' eye glances away from the forward roadway and the project team has tested this capability and system in previous studies (Donmez, Boyle, & Lee, 2007a, 2008, 2010).

The University of Iowa – National Advanced Driving Simulator (NADS) developed the software for the study and integrated the tasks so that they seamlessly worked with the driving simulator and eye tracker (for the UW study). For the occlusion study at the UW-Madison, a Ford Fusion vehicle simulator cab (Figure 4b) and the CogLens visual occlusion goggles (Figure 4a) were used for data collection. The occlusion goggles provided a brief “vision” interval to perform the text entry and reading tasks. This was followed by a brief “occlusion” interval to define and separate the visual access intervals. NADS integrated the occlusion goggles into the text reading and text entry tasks so that the shutter close and shutter open intervals were controlled according to the ISO 16673:2007 guidelines (i.e., 1.5-second vision: 1.5-second occlusion cycles, alternating continuously) and the resulting data (shutter open or close condition at any given frame) were logged with the task performance data (touchscreen manipulations, e.g., letter entry or “ENTER” pressing). The CogLens system was used instead of the PLATO system (as was used in the VRTC study) because CogLens can be integrated more easily with modern computer systems. Since occlusion time is comparable to eyes-off-road time as observed from a simulator study, the task performance from the occlusion study is comparable to that of a simulator study. Figure 4 b shows the experimental setting in the occlusion condition (left) and the way a participant would interact with the touchscreen (right).

(a)

(b)

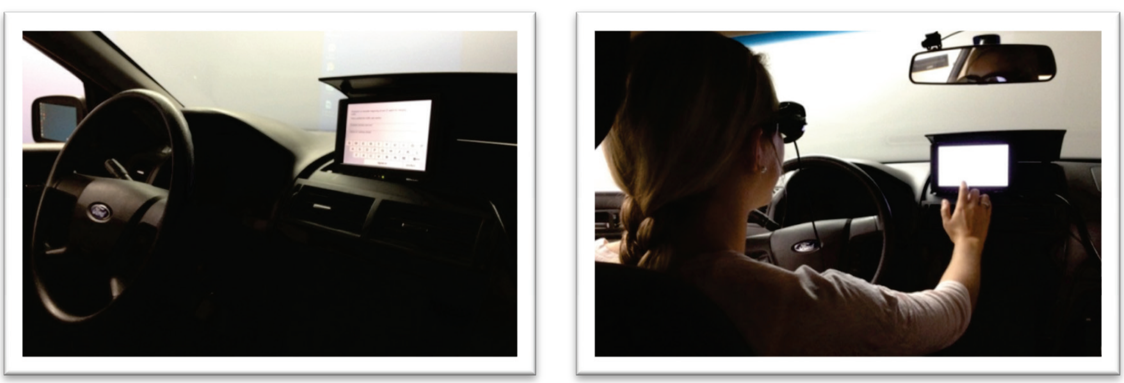


Figure 4. (a) CogLens visual occlusion goggles as shown at http://coglens.com/occlusion-glasses.html, and (b) the occlusion test experimental setting.

A 7-inch touchscreen display with QWERTY keyboard was used in both studies for text input and text reading tasks. The keyboard only contained keys that were needed for the tasks. This included capital alpha characters, backspace, enter, and space bar (i.e., no symbols, shift, or number pads). There was no type-ahead feature built into the keyboard. The touchscreen display was attached on the right side of the simulator dashboard at an appropriate location (Figure 5). Driver face video was recorded during data collection in the simulator study.

Figure 5. The NADS MiniSim with the 7-inch touchscreen display.

2.2.2 Driving Simulator Scenario

The driving scenario was designed by NADS, and followed the recommendations outlined in the proposed NHTSA Distraction Guidelines (Section X. VI.3). The simulated road (shown in Figure 6) included:

- Four lanes, undivided
- A solid double yellow line down the center, solid white lines on the outside edges, and dashed white lines separating the two lanes in the same direction
- Flat, straight road (no horizontal or vertical curves) and have a posted speed limit of 55 mph.

Figure 6. Example of simulator road (from NADS MiniSim).

All tests were performed while drivers were following a lead vehicle in the left lane of the simulated road. This is different than what was specified in the proposed NHTSA Guidelines where driving was specified to be in the right lane. It is unclear whether this difference may have impacted the outcome, but there were no lane changes performed during the study. Drivers were asked to maintain a travel speed of 50 mph, keep a 2-second headway from the lead vehicle, and drive safely. The scenarios did not include any lead vehicle braking events. The lead vehicle varied its speed based on a sinusoidal function defined by the programmers/simulator designers at NADS.

2.2.3 Independent Variables

The study used a mixed factorial, complete block design with 3 within-subject independent variables (Figure 7): task type (2 levels: Text Entry and Text Reading), text length (3 levels: Short, Medium and Long), and ambient text (2 levels: Present and Not Present). Based on this design, there were 12 different test conditions, with three replications for each test condition.

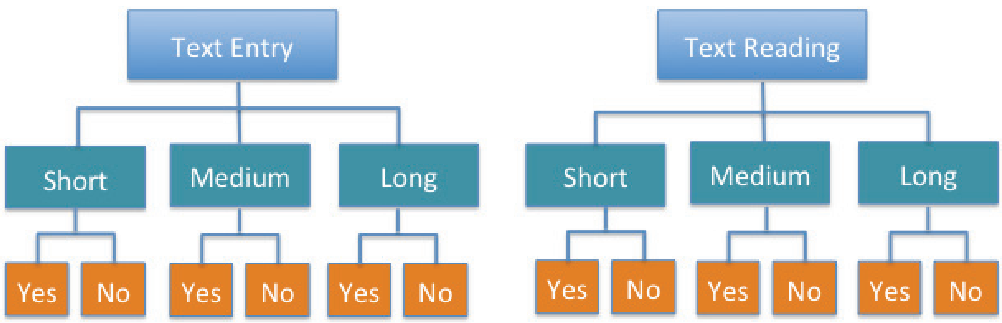


Figure 7. Experimental design with three independent variables: task type, text length, and ambient text.

Task type (2 levels). In line with the goals of the solicitation and the proposed NHTSA Distraction Guidelines, two distraction task types considered to interfere with a driver's ability to safety operate the vehicle were tested:

- Text entry: driver's ability to manually enter text using button or key presses in a single task.
- Text reading: driver's ability to read static (non-scrolling) text while driving.

Text length (3 levels). Short, medium, and long text lengths for each task type were examined. For the text entry, this is based on character strings that are similar to data entry for street names. The proposed NHTSA Distraction Guidelines (Section VI. H) indicate, "drivers should not input more than 6 button or key presses during the performance of a manual text entry (e.g., drafting text messages, keyboard-based text entry). This limit is based on an assumed driver EOR time of 2.0 seconds per button or key press and NHTSA's maximum permitted total EOR time for a task of 12.0 seconds." As a consequence, this study brackets the 6-key press guideline with substantially higher and lower values.

The text strings used for the text entry task were mostly street names found in road database files. However, other available lists (e.g., word game word lists) were also explored to create the text entry words. Table 3 shows some examples of the 4, 6, and 12 characters words that were used for the text entry task.

Table 3. Character Length for Text Entry

4 (Short)	**6 (Medium)**	**12 (Long)**
Main	Valley	Pennsylvania
Pike	Boston	Commonwealth
Lake	Nevada	Mountainside

The text reading condition was designed to include character strings that are typically observed on changeable message signs. As discussed in the Introduction, there are several studies by Dudek et al. (2007; 2006) that identify the proper display of information on CMS and were used for designing the task entry tasks. Table 4 provides examples of phrases that were used for the text reading task.

Table 4. Examples of Text Reading Task

Character Length	**Example Text**
Short (20 to 40 char)	Disabled vehicles next exit [26 char]
Medium (60 to 80)	There are two disabled vehicles on the right shoulder up ahead [62 char]
Long (120 to 140)	There are two disabled vehicles on the right shoulder up ahead move to the right most lane as quickly as possible [125 char]

The short character length for the text reading task was based on the proposed NHTSA guidelines of 30 characters. The long level of the text reading task was based on the reading time of text messages while driving. In a driving simulator study (Hoffman, Lee, & McGehee, 2006), drivers were asked to read multiline text messages. They were observed to use a series of 12 to 17 glances, each lasting 0.75 to 1.15 seconds, to read messages that were 8 to 9 lines long (or approximately four words for each glance). These results suggest that drivers can read up to 32 words (8 lines x 4 words/line), in approximately 16 seconds of off-road glance time (16 glances, each lasting 1 second) corresponding to 2 words per second of off-road glance time (a reading time of 10 characters per second of off road glance time). If each word consisted of 5 characters, then a message of 120 characters would produce the 12-second maximum off-road glance duration. Thus, a message of 120 characters would place the limit of the proposed NHTSA Distraction Guidelines at the maximum threshold of, "cumulative time spent glancing away from the roadway of 12 seconds" (Section I. C). This is also consistent with the 30-character limit that JAMA guideline stated when kanji characters are translated into English letters as discussed in section 1.1.3.3.

Ambient text (2 levels). There are many types of ambient text that can draw a driver's attention and this study will focus on one type with two levels: Yes or No

- For level NO: Participants would only see what they are entering or reading (Figure 8)
- For level YES: Ambient text is present during the text entry and text reading conditions (Figure 9)

TARGET text (for reading and entry) was displayed in a random row location, and was identified as the TARGETED TEXT within a box (Figure 9). The specific font used for the tasks was Arial narrow, with 1/8 inch height, displayed at a distance of about 2.3 feet away from driver's eye point, resulting in a visual angle of 0.26 degrees.

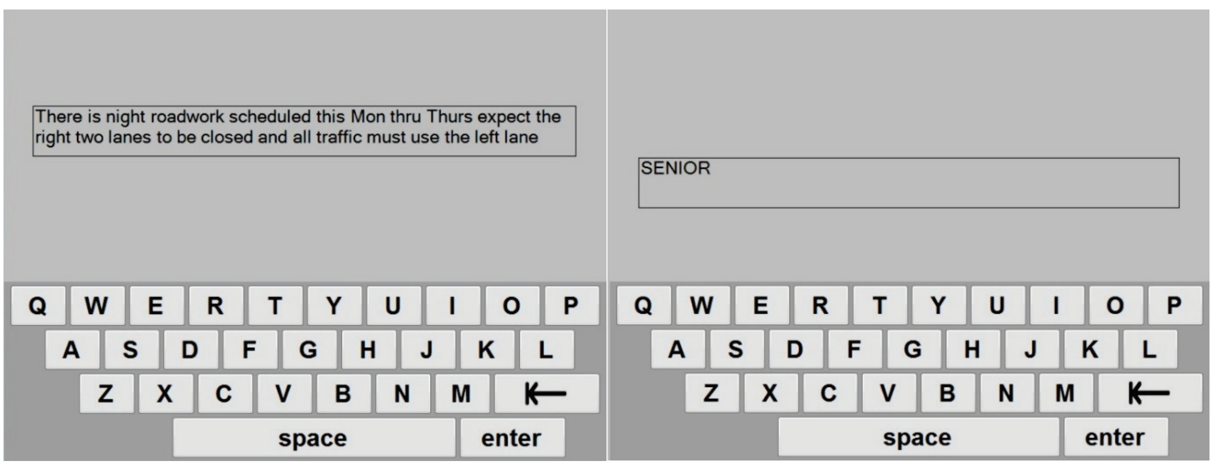


Figure 8. Examples of screens with NO ambient text for text reading (left) and text entry (right) tasks.

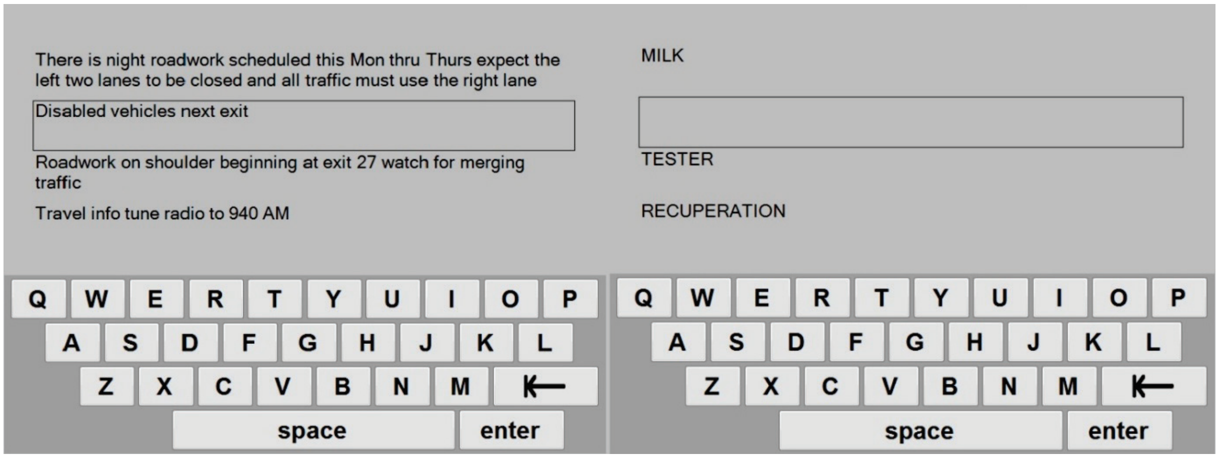


Figure 9. Examples of screens WITH ambient text for text reading (left) and text entry (right) tasks.

Participants never saw the same text entry word or text reading phrase more than once during the entire simulator/occlusion trials. That is, the text entry words and reading phrases were randomly selected and displayed without replacement. There were 24 words/24 phrases to select from for each text length condition (or a total of 72 words/72 phrases). The proposed NHTSA Distraction Guidelines stated one replication for each test condition during the simulator study and multiple replications for each condition during the occlusion study (Section X Option EGDS and Option OCC). ISO 16673 more specifically state five replications per test condition for the occlusions study. The current study used three replications per test condition to increase the test efficiency in the driving simulator and more importantly, allow direct comparisons with the occlusion study. In summary, each participant completed 36 different and randomly selected tasks (18 text reading and 18 text entry) across two simulator (or occlusion) trials. For each test condition, there was a total of 84 data points (28 participants x 3 replications) for the simulator (or occlusion) trials.

2.2.4 Task Content and Order

To allow direct comparisons between the simulator and occlusion experiments, each study included a static condition (as recommended in the ISO/CD 16673). Each participant completed 2 static and 2 other trials (driving simulator or occlusion). Inclusion of static trials is not stated in the proposed NHTSA Guidelines requirements but it provides an indicator of the driver's ability to read and enter text, which could be used as a covariate in future analysis that compare the performance from the simulator and occlusion trials. Such a covariate would enhance the statistical power of the study considerably and, more importantly, it could provide insights on any differences observed in the simulator and occlusion test results. The text reading and enter tasks was the same for the static trials, simulator, and occlusion) trials (i.e., 36 randomly selected tasks with 18 text entry and 18 text reading tasks across two trials).

To minimize carry-over and training effects, the experimental order followed the recommended experimental design outlined in ISO/CD 16673:

Driving Simulator Study (UW)

- Trial Order 1: Driving Simulator → Static
- Trial Order 2: Static → Driving Simulator

Occlusion Study (UW-Madison)

- Trial Order 1: Occlusion → Static
- Trial Order 2: Static → Occlusion

For both studies, 50 percent of participants received trial order one and the other 50 percent received order two. The orders were randomly assigned to participants. The number of participants that received each order by age group is shown in Table 5 (same for both driving simulator and occlusion studies).

Table 5. Trial Order by Age Group

Trial Order	18-24	25-39	40-54	55-75
Order 1	4	3	4	3
Order 2	3	4	3	4

2.2.5 Procedures

The experimental procedures for the driving simulator and occlusion study were executed similarly and followed IRB protocols approved by the UW (IRB No. 42893) and the UW-Madison (IRB No. SE-2012-0247). Potential participants were screened via telephone to determine if they were qualified for the study. If the participant met all inclusion criteria and agreed to participate in this study, the research team would arrange an experiment time. Upon arrival, participants were asked to show their driver's licenses to confirm that they were valid. The experimenter verbally reviewed the informed consent form (ICF), and obtained participants' written consent. All participants were provided with a copy of the signed ICF. The participants filled out a payment form and completed a questionnaire that covered some general questions about their driving, demographics, vision, hearing, and motion sickness related to use a driving simulator.

The participants received the appropriate training (driving simulator at the UW and occlusion goggles in UW-Madison), which included test instruction for performing the text entry and reading tasks. For the simulator study, the participants had their eyes calibrated for the eye tracker followed by a practice trial. In the practice trial, the participant would drive in the simulator without using the touchscreen to get familiar with driving at the speed limit, and the use of the brake and gas pedals, and the steering wheel. The practice text entry and reading tasks would start approximately 5 minutes into the practice trial. The practice tasks included three entry and three reading tasks with the length of short, medium and long text length. After the practice trial, participants were asked to complete a questionnaire about how they feel in the simulator to ensure that there was a small likelihood for simulator sickness.

For the occlusion study, there were a set of three practice tasks in the static condition (without goggles) and a set of three practice tasks in the occlusion condition. Any questions that the participants might have about the tasks were addressed during this time.

There were four trials in both the simulator and occlusion study, with two static trials and two the simulator/occlusion trials. The trial order for each participant was randomly assigned before his/her arrival. For each trial, participants were given 18 tasks related to text reading (9 phrases) and text entry (9 words), in a random order (without replacement). There were a total of 72 tasks (4 trials x 18 tasks) encountered when all four trials were completed.

The simulator study lasted approximately 1 to 1.5 hours, including 5-minute breaks between trials. In the simulator trials, the tasks would automatically start around 40 seconds to 1 minute into the drive. Participants were asked to keep driving for 5 minutes after they completed all the tasks. The static trials were approximately 8 to 10 minutes in duration, and the simulator trials were approximately 15 minutes. The occlusion study lasted approximately 1 hour, which each trial approximately 8 to 15 minutes long. Participants were given 5-minute breaks in between each trial.

For the text entry task, the computer would provide an auditory cue with a word to the participant. The participant would need to enter the keystrokes corresponding to the word just heard using the QWERTY touchscreen keyboard. Participants would need to press the "enter" key once they have completed the sequence of keystrokes.

For the text reading task, the participant would hear a beep from the computer indicating that a phrase was to be read within a box on the touchscreen. They were instructed to press the "enter" key when they have completed and comprehended the phrase displayed. The computer would then read a statement related to the phrase that the participants just read on the screen. The participants would need to select either "True" or "False" to decide whether the statement was correct (Figure 10). For both text entry and reading tasks, participants were instructed to provide their best guess if they did not know the answer. The protocols for the text entry and reading tasks are shown in Table 6.

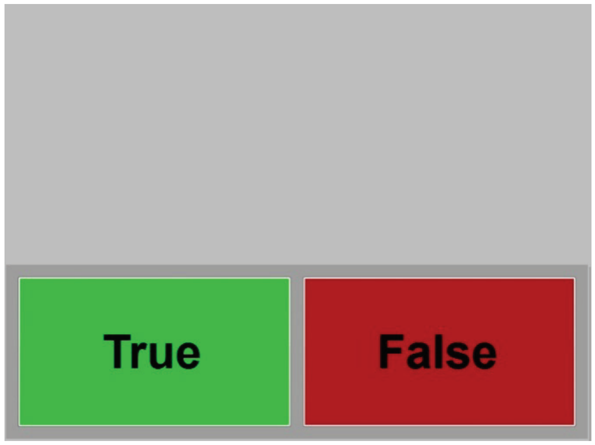


Figure 10. Touchscreen interface for true/false selection in reading tasks.

Table 6. Protocol for the Text Entry and Text Reading Tasks

Text Entry	Text Reading
• The computer says a word • User enters keystrokes for word • User presses ENTER when complete	• The computer displays a phrase • User presses ENTER after comprehending • The computer reads a statement • User presses True or False

2.2.6 Power Analysis for ANOVA Tests

Ranney et al. (2012) showed that a sample size of 80 participants should be used to obtain a power of 0.8 and maintain adequate control of Type II error when performing the conformance/nonconformance tests. The proposed NHTSA Visual Manual Guidelines recommend using 24 participants. To increase the power of our criteria evaluation tests, the study involved 28 participants in the driving simulator, 28 different participants in the occlusion trials, and a combined total of 56 participants for the static trials.

A post hoc power analysis was conducted (G*Power 3.1.1) to assess the expected statistical power for repeated-measures ANOVA with three within-subject variables (2 task types × 3 text length × 2 ambient text = 12 conditions) and two between-subject variables (2 genders × 4 age groups = 8 conditions). Table 7 shows the expected statistical power to detect a small ($f = 0.1$), medium ($f = 0.3$) and large ($f = 0.5$) effect on each of the 12 within-subject test conditions (i.e., three-way interaction effect) when having one, two, and three replications on each test condition. For this study, with three replications per test condition, there was expected to be enough statistical power to detect any medium to large effects, and moderately large power to detect small effects in each condition.

Table 7. Power Analysis for ANOVA Tests

	Effect size		
No. of replications/condition	Small ($f = 0.1$)	Medium ($f = 0.3$)	Large ($f = 0.5$)
1	0.33	0.99	1.00
2	0.50	1.00	1.00
3	0.63	1.00	1.00

The calculation is based on assuming the correlation among repeated measures (ρ) is 0.5 and nonsphericity correction ε is 1; Noncentrality Parameter $\lambda = Nmf^2 / (1 - \rho)$, where N (total number of subjects) = 28, m (number of repeated measures) = 12, 24, and 36, respectively (see http://www.psycho.uni-duesseldorf.de/aap/projects/gpower/gpower-tutorial.pdf).

2.3 DATA REDUCTION

For the driving simulator study, data collected from the eye tracker, text entry and reading tasks were merged for data reduction and analysis. The eye tracking data include the time-stamped eye glance location and the corresponding XYZ coordinates (three dimensional world). The text task data included the specific test conditions, the time-stamped input stimuli, participant keystroke commands, and the time-stamped final participant entry. The eye tracking data and text task data were synchronized using the time stamps. The starting and ending points of each task were identified from these variables, and matched with the eye tracking data so that EOR durations can be calculated. All data were recorded at 60 Hz.

Eye tracking data often includes inaccuracies that must be addressed before subsequent analysis. Therefore, the eye tracking data were manually verified using the driver face videos recorded during the simulator trials. The driver face videos were recorded at the same rate as the eye tracking and driving simulator and all was time-stamped using the simulator. When the eye tracker failed to detect an eye movement or showed unusual glance locations, the researcher would select the participant's actual eye locations (e.g., touchscreen, center roadway) according to the video using timestamps. The XYZ coordinates recorded by the eye tracker were also used to verify the eye locations.

3. RESULTS

The findings of the simulator and occlusion study are discussed in this section with respect to the five acceptance criteria outlined in the proposed NHTSA Distraction Guidelines. The NHTSA Distraction Guidelines indicate that 85 percent of participants must conform to the acceptance criteria. The original acceptance criteria were based on a sample size of 24 and hence, were rounded to 21 as needing to conform. In this study, 24 out of 28 participants have to conform for criteria 2, 3, 4, and 5.

Descriptive statistics are presented in this section, along with density, mean, and error bar plots by task type and text length. The density plots show the shapes of distributions and locations of percentiles of interest, whereas the mean and error bar plots show the summarized responses by age groups. Repeated measures ANOVA models were performed when appropriate to examine the significance of text type, text length and ambient text. Each test condition had three replications and the outcomes are discussed at the individual task level and also at the participant level. The forthcoming discussions demonstrate that there could be different conclusions regarding the acceptance criteria depending on the level of data aggregation.

3.1 DRIVING SIMULATOR STUDY

3.1.1 Demographics

There were 13 females and 15 males in the simulator study. The mean age of the four age groups are 21.1 (*SD* = 0.4, range: 21-22), 28.7 (*SD* = 4.3, range: 26-38), 48.6 (*SD* = 3.1, range: 45-54), and 59.7 (*SD* = 3.0, range: 56-64) years old. All participants had at least some college education. Among these participants, 11 (39.3%) had 4-year college degree, and 7 (25%) had master's or higher degrees.

The average age that participants indicated that they started to drive was 15.5 years old (*SD* = 1.35 years, range: 13–20 years) and obtained their first driver's license at the average age of 16.4 years old (*SD* = 0.97 years, range: 15.5–20 years). Fourteen participants reported that they drove at least once daily, 13 participants reported that they drove at least once weekly, and only one participant from the youngest group reported to drive less than once weekly. No participants had any crashes in the last year, with three participants reported that they had one crash in the last three years. Two participants had one moving violation in the last year, and one of them had three moving violations in the past three years.

3.1.2 Task Duration for Simulator Condition

Figure 11 shows the density of task durations under each test condition (each curve includes 168 data points: 28 participants × 3 replications × 2 levels of ambient text). The variation of task duration is very different across task type and text length. Text reading appears to have less variation than text entry. For text reading and text entry, there was a greater spread in time to complete tasks as the number of characters increased.

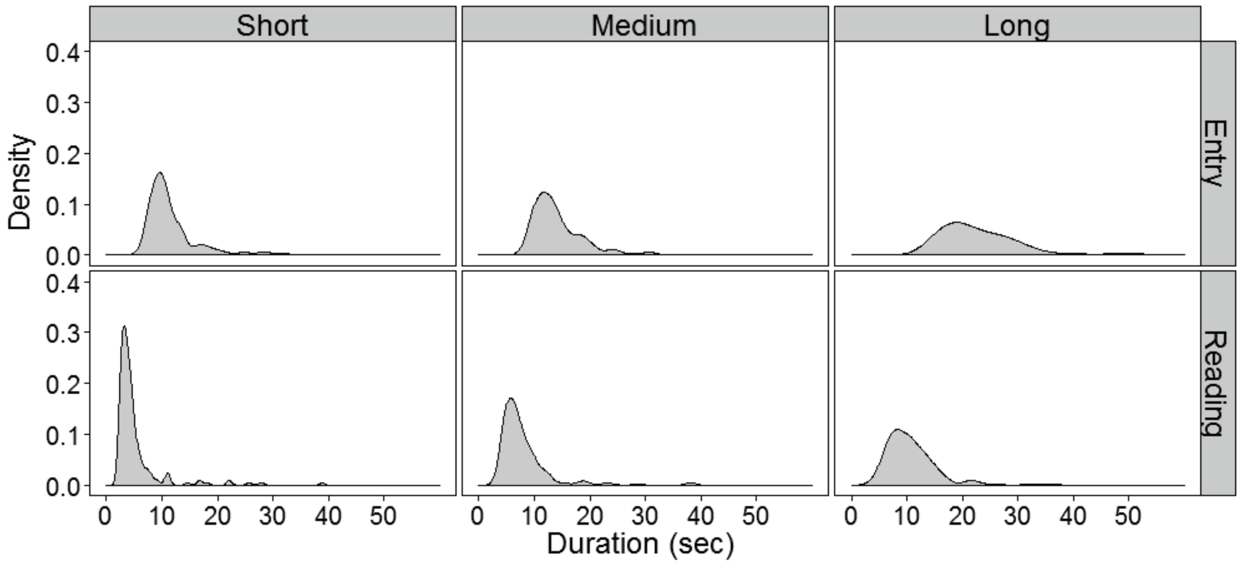


Figure 11. Density plot of task duration in simulator condition by task type and length (at the task level).

The task duration was aggregated on participant level by taking average of the three replications on each test condition. The mean task duration with inter-subject standard deviation for each test condition are shown in Figure 12 by age group. The solid dots show the averaged task duration for each participant. The task duration increases as the text length increases, regardless of age group, task type, or ambient text. The youngest age group appears to have the shortest mean task duration under the text reading condition when compared to other age groups. Ambient text appears to slightly increase the task duration for text reading, but not for text entry. The mean task duration for long text entry is over 23 seconds, while the mean for short and medium text entry are approximately12 and 14 seconds, respectively (Table 8). The task durations for text reading are much shorter than for text entry. The short reading without ambient text has the shortest mean task duration and smallest inter-subject variations (M = 3.95s, inter-subject SD = 1.16s).

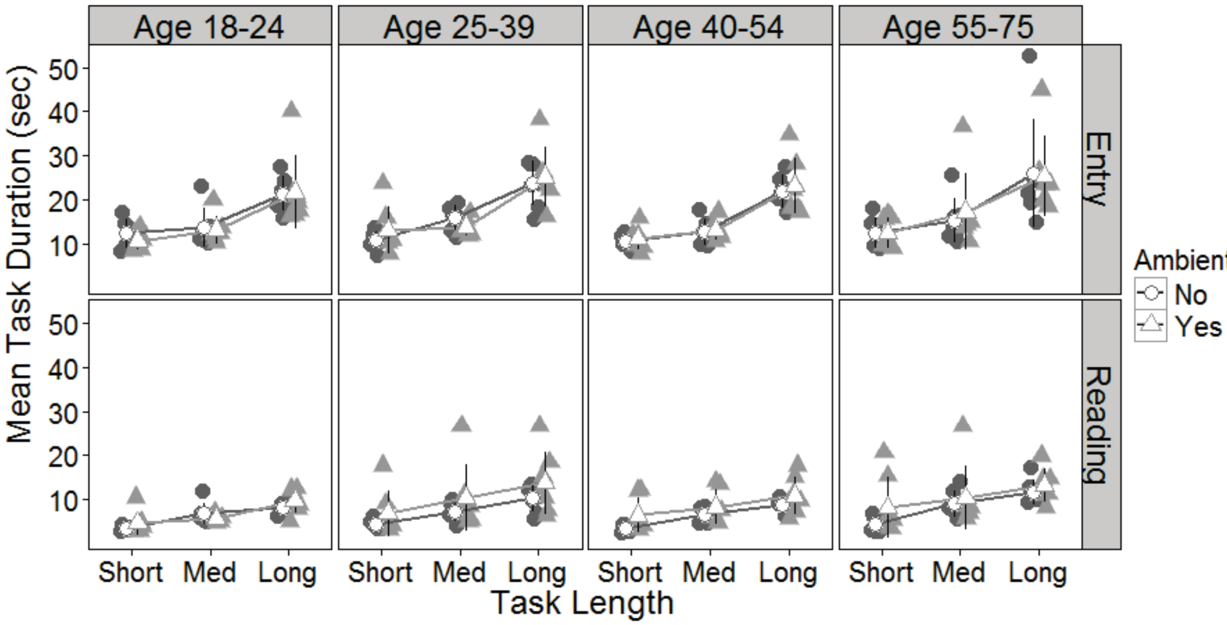


Figure 12. Mean task durations in simulator trials by age, for each task condition (at the participant level).

Table 8. Task Duration for Simulator Trials

	Length	Ambient	**Mean** (s)	**± 95% CI of mean** (s)	**Inter-subj *SD*** (s)
Entry	Short (4 char)	Yes	11.85	[10.72, 12.98]	3.54
		No	11.52	[10.62, 12.42]	2.64
	Med (6 char)	Yes	14.34	[13.23, 15.46]	4.89
		No	14.38	[13.50, 15.26]	3.92
	Long (12 char)	Yes	23.82	[22.23, 25.41]	7.43
		No	23.27	[21.93, 24.61]	7.02
Reading	Short (20 - 40 char)	Yes	6.54	[5.36, 7.72]	4.84
		No	3.95	[3.35, 4.56]	1.16
	Med (60 - 80 char)	Yes	8.66	[7.65, 9.67]	5.62
		No	7.39	[6.47, 8.32]	2.42
	Long (120 - 140 char)	Yes	11.86	[10.80, 12.91]	4.77
		No	9.64	[8.89, 10.38]	2.55

Note: The 95% CI of mean is calculated based on intra-subject standard error of mean

3.1.3 Criterion 1: Percentage of Long Eyes-Off-Road (EOR) (≥2sec)

The conformance criteria outlined in the proposed NHTSA Distraction Guidelines state that for at least 85 percent of the test participants, 85 percent of individual glance durations should be less than 2.0 seconds. Since we used 28 participants in this study, there should be at least 24 conforming to the criterion. In this study, an individual glance longer than or equal to 2.0 seconds is referred to as long EOR. For each participant, the percentage of long EOR for each task (or replication) is calculated as:

$$\text{\% Long EOR in task } i = \frac{\textit{No. of EOR} \geq 2 \text{ sec in task } i}{\textit{Total No. EOR} \text{ in task } i} \times 100\%$$

For example, if there are four glances in a single task and two of them were longer than 2 seconds, then the percent Long EOR in this task would be 50 percent. Performance on a single task conforms to criterion 1 when "*% Long EOR*" is less than 15 percent. Because each test condition was repeated three times for each participant, the evaluation for each test condition is based on the proportion of replications. Specifically, Criterion 1 is examined based on three yardsticks: conforms if the participant is in conformance in (1) all three replications, (2) at least two out of three replications, or (3) at least one out of three replications.

Figure 13 shows the histogram of the percentage of long EOR under each test condition (each cell includes 28 participants × 3 replications = 74 data points). The vertical bars show the number of tasks that have a certain percentage of long EOR. The dotted vertical lines show the 15 percent acceptance criterion threshold, and the bars on the left side of the dotted line conform to the threshold. Many individual tasks were not in conformance with the acceptance criterion of 15 percent or less for long EORs; and this was observed in all test conditions. It can be seen that regardless of task type and ambient text, the number of individual tasks that have 0 percent long EOR (i.e., none of the EORs in the task is longer or equal to 2 seconds) decreases as the text length increases. More tasks have 0 percent long EOR in text reading than in text entry conditions. For example, about 70 percent of short text reading tasks do not have any long EOR, while roughly 40 percent of short text entry does not have long EOR.

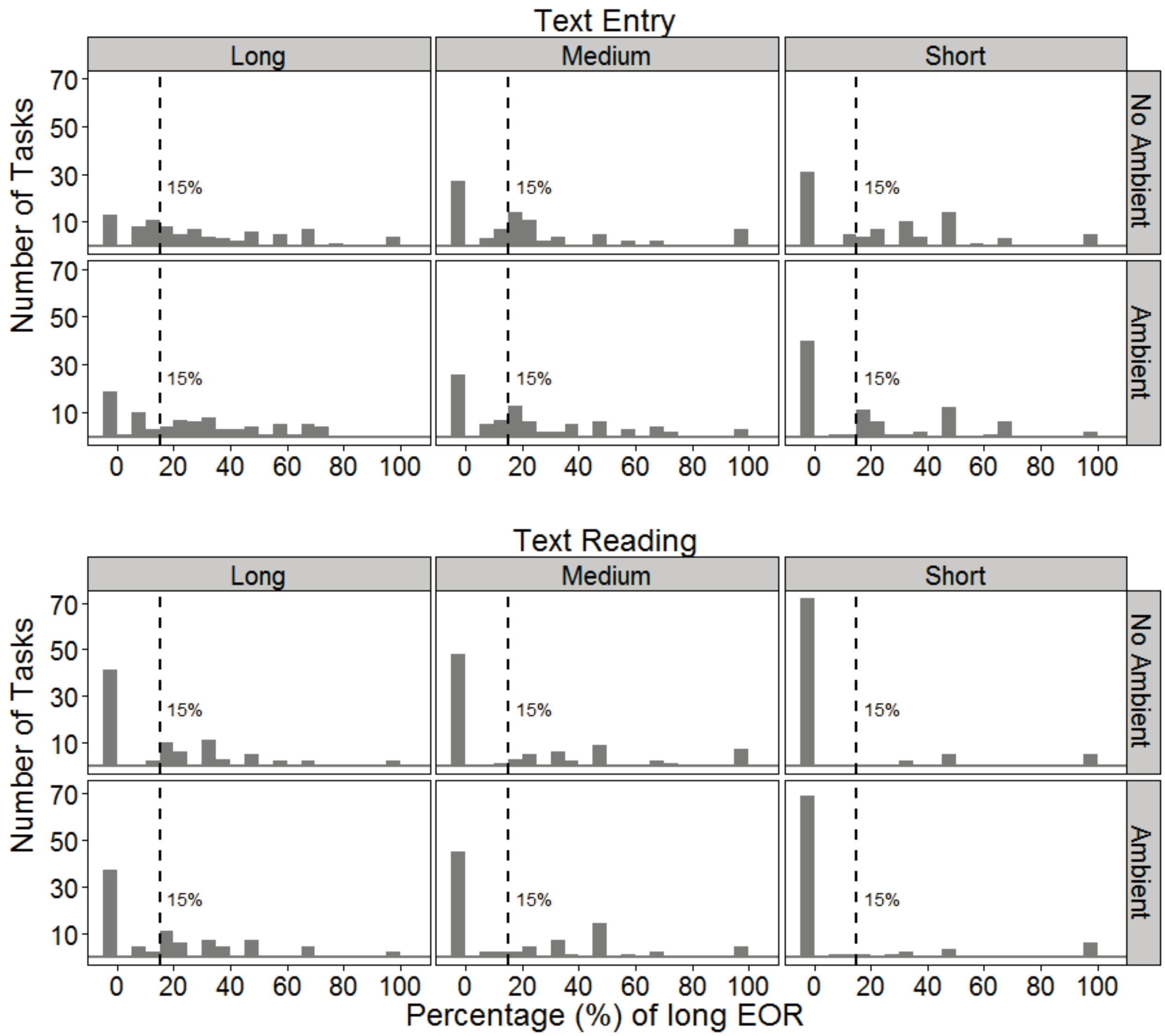


Figure 13. Histogram of the percentage of long EOR for each task condition (at the task level).

Table 9 shows the results aggregated by both task level and participant level. On the task level, no more than 50 percent of the tasks conformed to the acceptance criterion under each of the six text entry tests. On the other hand, at least 50 percent of tasks conformed to the acceptance criterion under each of the six text reading tests. The number of tasks conformed to the criterion largely decreased from short to medium and long text reading conditions, but only slightly decreased from short to long text entry conditions. Additionally, ambient text does not seem to have an effect on the conformance.

For text entry, no more than 25 percent of the participants (7 or less) were able to conform to this criterion with all three replications (Table 9). Depending on the interaction level, compliance was not met by 25 to 46 percent of participants for any replication. Therefore, none of the text entry conditions conformed to the test criteria. For the text reading condition, approximately 60 percent of participants were in conformance in all 3 replications and more than 90 percent of participants were in conformance in at least 2 out of 3 replications in the short reading condition. For medium and long text reading, only about 30 percent of participants were in conformance in all 3 replications, and about 50 percent were in conformance in 2 out of 3 replications. For the short and medium text reading, at least 85 percent of participants were in conformance for at least 1 out of 3 replications (as shaded in Table 9).

Table 9. Number of Participants and Tasks That Conformed With the Acceptance Criterion 1

Task Type	Length	Ambient	Number out of 28 participants (and %)			Tasks comply (out of 84 tasks)
			3 reps comply	2+ reps comply	1+ reps comply	
Entry	Short	Yes	7 (25.0)	14 (50.0)	21 (75.0)	42 (50.0)
		No	5 (17.9)	12 (42.9)	19 (67.9)	36 (42.9)
	Med	Yes	7 (25.0)	14 (50.0)	17 (60.7)	38 (45.2)
		No	7 (25.0)	12 (42.9)	18 (64.3)	37 (44.0)
	Long	Yes	7 (25.0)	11 (39.3)	15 (53.6)	33 (39.3)
		No	5 (17.9)	10 (35.7)	17 (60.7)	32 (38.1)
Reading	Short	Yes	18 (64.3)	26 (92.8)	27 (96.4)	71 (84.5)
		No	17 (60.7)	27 (96.4)	28 (100)	72 (85.7)
	Med	Yes	10 (35.7)	15 (53.6)	24 (85.7)	49 (58.3)
		No	9 (32.1)	16 (57.1)	24 (85.7)	49 (58.3)
	Long	Yes	9 (32.1)	13 (46.4)	21 (75.0)	43 (51.2)
		No	5 (17.9)	15 (35.7)	23 (82.1)	43 (51.2)

Note: Shaded cells show the test condition that *conformed to* the acceptance criterion

In conclusion, the text reading produced smaller percentage of EOR ≥ 2 sec than text entry. Additionally, shorter tasks produce smaller percentage of EOR ≥2 sec than longer ones. This effect of text length appears to be smaller for text entry than for reading. The number of replications in each test condition and the acceptable proportion of nonconformance in these replications can impact the conformance rate. Based on the current testing protocol, the only test conditions that can be considered to conform to the proposed NHTSA Distraction Guidelines is the short and medium text reading cases with or without ambient text (*when a test condition is considered in conformance for a participant if he/she complies in at least one out the three replications)*, or the short text reading cases with or without ambient text (*when a test condition is considered in conformance for a participant if he/she complies in at least two out the three replications)*.

3.1.4 Criterion 2: Mean Glance Duration

The acceptance criterion based on the proposed NHTSA Distraction Guidelines is that 85 percent of participants' MGD should be less than 2.0 seconds. This study used 28 participants that this criterion rounds to approximately 24 participants needing to have MGDs less than 2.0 seconds. For each participant, the MGD for a single replication is defined as:

$$MGD\ in\ task_i = \frac{\sum_{j=1}^{n_i} EOR\ Duration_{j_i}}{n_i},$$

where *EOR Duration* $_{ji}$ is the jth EOR in task i, and n_i is the total number of EOR in task i.

Since there are three replications of each test condition, the MGD is calculated as the averaged mean glance durations per test condition:

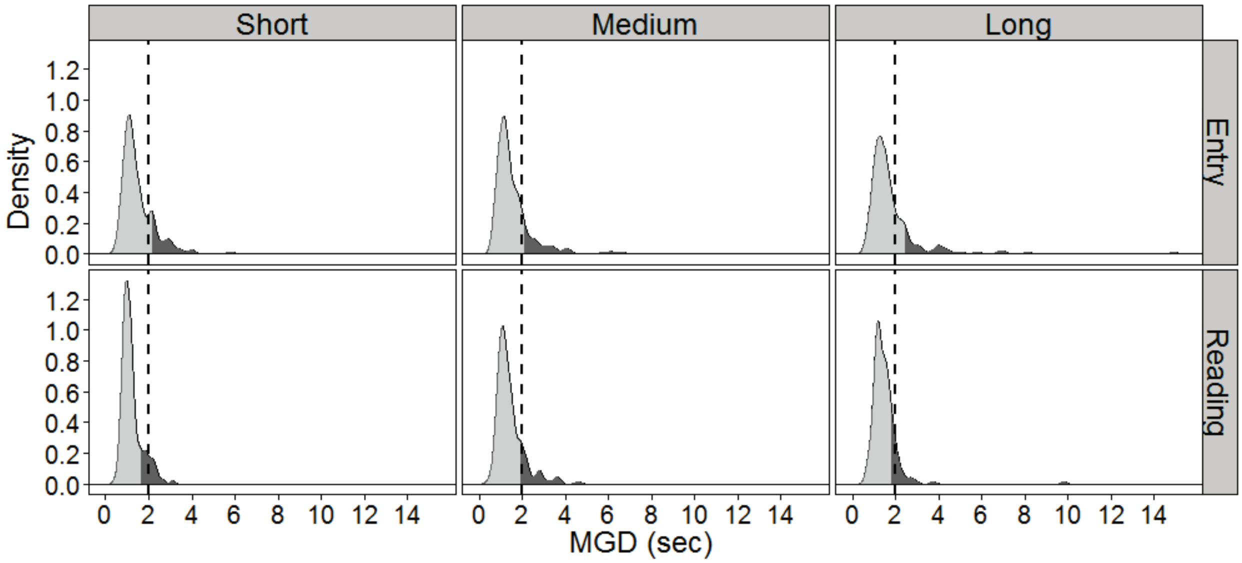
$$MGD = \frac{\sum_{i=1}^{3} MGD\ in\ task_i}{3}$$

This averaged MGD is then used to evaluate each participant's conformance with the criteria on a certain test. Specifically, a test condition is considered in conformance for a participant if the MGD < 2 seconds. In addition, the test condition is considered to conform with criterion 2 if at least 24 out of the total 28 participants (85%) were in conformance in this test (or equivalently, no more than 4 out of the 28 participants (15%) does not conform).

3.1.4.1 Descriptive Analysis

Figure 14 shows the density plots of MGD for individual tasks separated by task type and length (N = 28 participants × 2 levels of ambient text × 3 replications =168 data points). The black vertical dot line in each plot shows the acceptance criterion of 2 seconds. The light gray area shows the proportion of the distribution that falls below the 85th percentile and the dark gray area shows the upper 15 percent of the distribution. Therefore, more than 15 percent of all tasks have MGD ≥ 2 seconds if the dotted line falls in the light gray area. This can be seen clearly as the case for all text entry conditions. For all text reading, slightly more than 85 percent of the tasks conformed with the 2-second criterion.

The variation on MGD is slightly larger for text entry than text reading, and increases slightly comparing the short to long text length conditions (Figure 14). Additionally, the distributions of the long text entry and long text reading condition have longer tails on the right than the other conditions. The largest MGD observed for the long text entry is about 14 seconds, and for the long text reading, it is 10 seconds. The extreme glances that occurred during the long text entry and reading tasks may be due to the high demands of the tasks, as drivers need to enter words with as many as 12 characters, or read sentences that could be as long as 130 words. Some long length tasks were completed with a single long glance. It is rare to observe such extreme glance behavior. However, Horrey & Wickens (2007) did observe long single glance durations of 6 seconds for visual and cognitive tasks that lasted no more than 9.5 seconds.



Note: light gray shows proportion that falls below the 85th percentile; dark gray shows upper 15 percent

Figure 14. Density Plot of MGD by task type and length (at the task level).

The MGD on each of the 12 test conditions is shown in Figure 15, separated by each age group. The error bar shows the inter-subject standard deviation on the MGD, and the solid dots show the averaged MGD for each participant. The MGD is slightly smaller for text reading than text entry under each sub-test condition. There was also a slight increasing trend on MGD from short to long condition, but this trend was more obvious for text entry than for text reading. The horizontal dotted line shows the 2-second criterion. For text reading, the MGDs for most participants were below this 2-second limit. For text entry, approximately half of the participants in the three youngest age groups exceeded the limit for MGDs. Experimental observations showed that participants who had longer glances also entered more letters in a single chunk, which may provide an explanation for the longer MGD for text entry tasks. Additionally, older drivers appeared to be most cautious, as many had glances less than 2-second in order to maintain safe driving.

There also appears to be some individual differences on the MGD. Specifically, some participants have much larger MGD than others, which caused somewhat large inter-subject variations. For example, one participant in the 40–54 age group had an MGD that was over 8 seconds, and one participant in the 18–24 age group had an MGD around 6 seconds. The inter-subject variation appeared to be larger for text entry than text reading, and increased with the text length. However, the inter-subject variation is relatively small for the oldest participants (55 to 75 year olds) when compared to the other age groups under all test conditions.

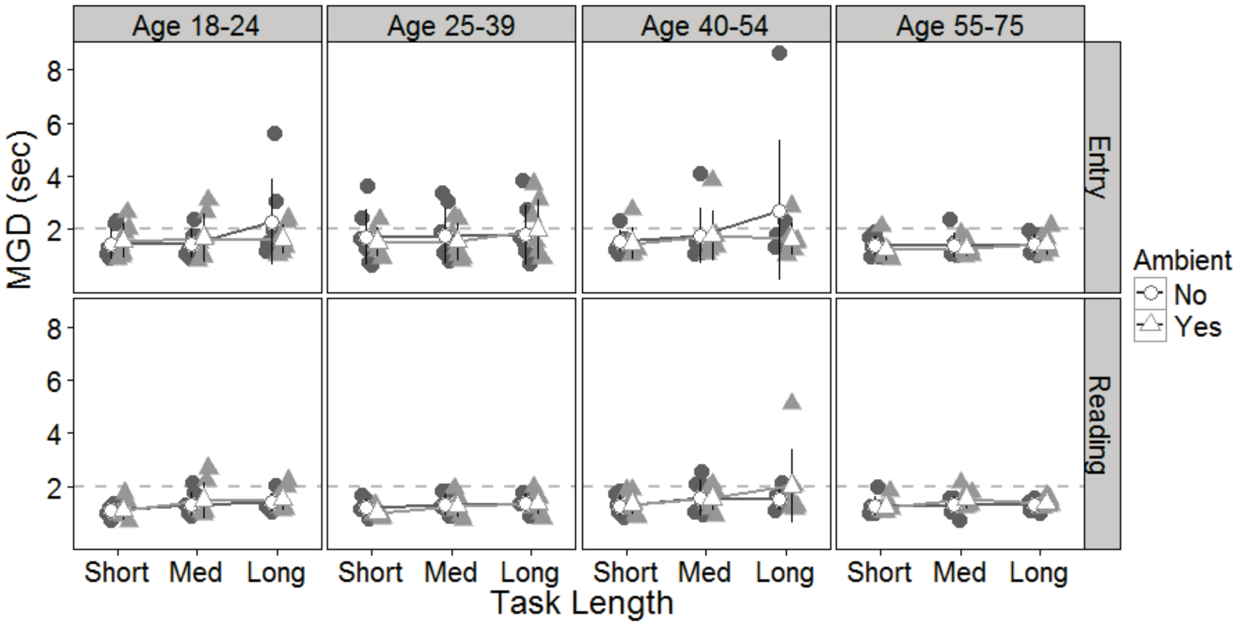


Figure 15. MGD by age group and each task condition (at the participant level).

Table 10 shows the number of participants not conformance with the acceptance criterion 2. As observed, none of the text entry conditions were in conformance with criterion 2 since more than four participants had MGDs greater than 2 seconds, and over 20 percent exceeded the 2-second threshold in the long text entry conditions. There were no differences in the number of participants (n=23) that were in conformance in the short and medium length text entry. Further examination of these participants revealed that most participants who had MGDs longer than 2 seconds in the short text entry also had long MGDs in all or most medium and long text entry conditions. This may indicate that, for drivers who tend to have longer glances off the road, shorter text entry may decrease their MGDs slightly but may not be low enough to bring their EOR durations to the safe range.

For text reading, the long reading without ambient text, and short reading with and without ambient text, conformed to the acceptance criterion as no more than two participants exceeded the 2-second threshold. Other conditions in text reading marginally conformed to criterion 2 (4 participants were not in conformance). Participants who were not in conformance in most of the text entry conditions did not necessary not conform in the reading conditions, which indicate different eye glance behaviors between text entry and reading.

If the acceptance criterion was determined based on task level instead of participant level, then 72 tasks (or 85%) or more must conform to criterion 2. In this case, only the medium text reading length without ambient text marginally conformed to criterion 2 with 72 out of 84 tasks being in conformance; all other text reading conditions conformed to the criterion. This demonstrates the impact on the evaluation given repeated measures and different levels of data aggregation.

Table 10. Mean Glance Duration

Task Type	Length	Ambient	Mean (s)	± 95% CI of mean (s)	Inter-subj *SD* (s)	No. subj out of 28 (NC) (and %)	No. tasks out of 84 (NC) (and %)
Entry	Short	Yes	1.44	[1.32, 1.56]	0.54	5 (17.9)	15 (17.9)
		No	1.52	[1.38, 1.66]	0.63	5 (17.9)	20 (23.8)
	Med	Yes	1.57	[1.42, 1.72]	0.73	5 (17.9)	13 (15.5)
		No	1.59	[1.45, 1.72]	0.80	5 (17.9)	16 (19.0)
	Long	Yes	1.67	[1.54, 1.81]	0.71	7 (25.0)	19 (22.6)
		No	2.06	[1.69, 2.42]	1.63	6 (21.4)	21 (25.0)
Reading	Short	Yes	1.17	[1.04, 1.30]	0.33	0 (0.0)	6 (7.1)
		No	1.21	[1.08, 1.34]	0.31	1 (3.6)	7 (8.3)
	Med	Yes	1.45	[1.32, 1.59]	0.48	4 (14.3)	10 (11.9)
		No	1.38	[1.25, 1.51]	0.46	4 (14.3)	12 (14.3)
	Long	Yes	1.57	[1.38, 1.75]	0.77	4 (14.3)	11 (13.1)
		No	1.39	[1.29, 1.50]	0.31	2 (7.1)	6 (7.1)

Note: The shaded cells shows the test conditions that *conformed with* the acceptance criterion.
The 95 percent CI of mean is calculated based on intra-subject standard error of mean.

3.1.4.2 ANOVA for Mean Glance Duration

A linear mixed model (R. 2.12.1) was used to perform the repeated measure ANOVA on the response variable, MGD. MGD was also log transformed to meet the assumptions associated with a regression model. The final model was chosen based on the lowest AIC (Akaike information criterion) value, which assesses the relative goodness of fit of the model (Table 11). The within-subject factors task type and length have significant main and interaction effects. Specifically, text entry has larger MGD than text reading. For text entry, there is a larger increase on MGD from medium to long text length than from short to medium text length (Figure 16). For text reading, the increase on MGD from short to medium text length is larger than from medium to long length. Additionally, there is a significant interaction effect between task type and ambient text. The MGD is smaller when there is ambient text than without ambient text for text entry, whereas the opposite is observed for text reading. In other words, it appears to be easier for drivers to ignore clutter in the text entry condition, but more difficult to ignore clutter when trying to read sentences.

There is also a significant two-way interaction effect on gender and age group, as well as a three-way interaction effect between task type, gender and age group. However, given the small sample size in each gender and age group and thus the large confidence interval for mean (lower left plot in Figure 16), the effects of gender and age need to be interpreted with caution.

Table 11. ANOVA Results on MGD

	Num df	Den df	F-value	p-value
(Intercept)	1	962	63.75	<.0001
Task Type	**1**	**962**	**14.26**	**<.0001**
Gender	1	20	0.92	0.35
Age Group	3	20	0.71	0.56
Length	**2**	**962**	**27.78**	**<.0001**
Ambient	1	962	0.30	0.58
Gender * Age Group	**3**	**20**	**5.26**	**0.01**
Length*Ambient	2	962	0.94	0.39
Task Type *Gender	1	962	0.47	0.49
Task Type *Age Group	3	962	1.70	0.17
Task Type * Length	**2**	**962**	**3.00**	**0.05**
Task Type * Ambient	**1**	**962**	**4.61**	**0.03**
Task Type * Gender * Age Group	**3**	**962**	**8.17**	**<.0001**
Type * Length * Ambient	2	962	1.45	0.24

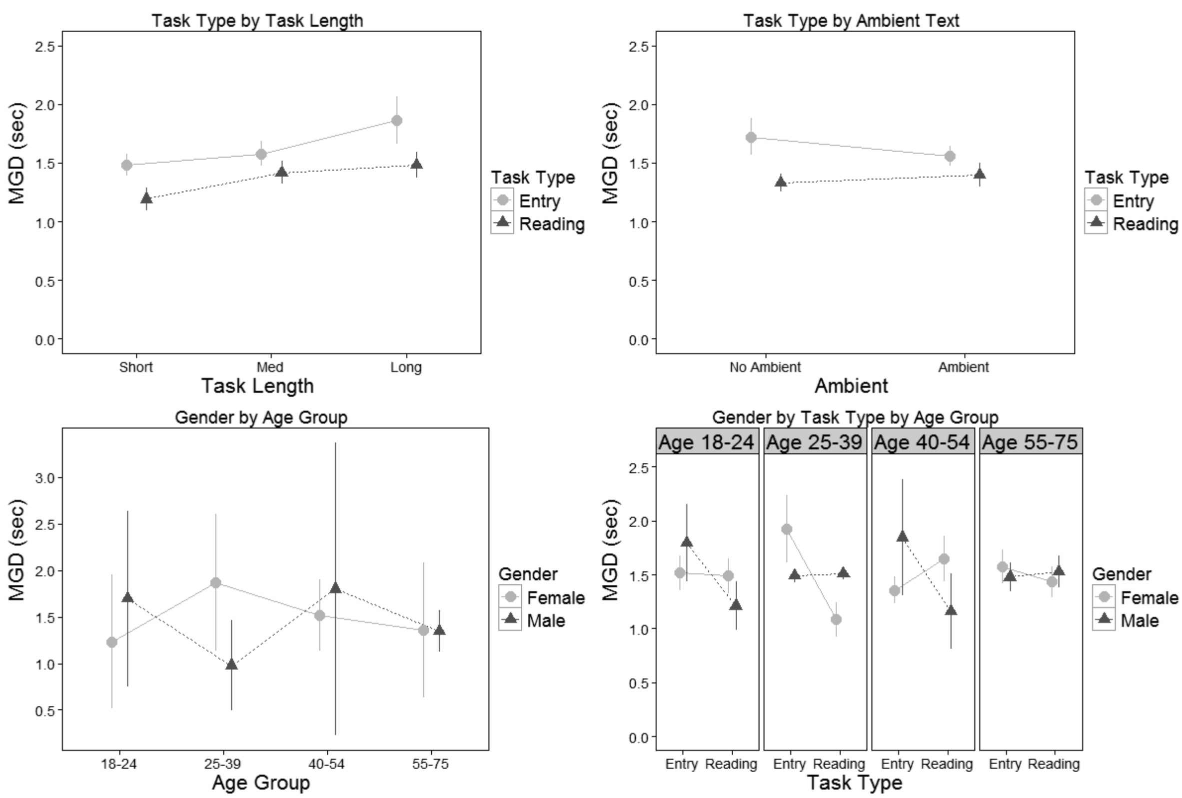


Figure 16. Interaction plots for MGD.

Note: the error bar in the Gender by Age Group plot (lower left) shows the 95 percent CI for mean based on inter-subject standard error; the error bars in all other three plots show the 95 percent CI for mean based on intra-subject standard error.

3.1.5 Criterion 3: Total Eyes-Off-Road Time

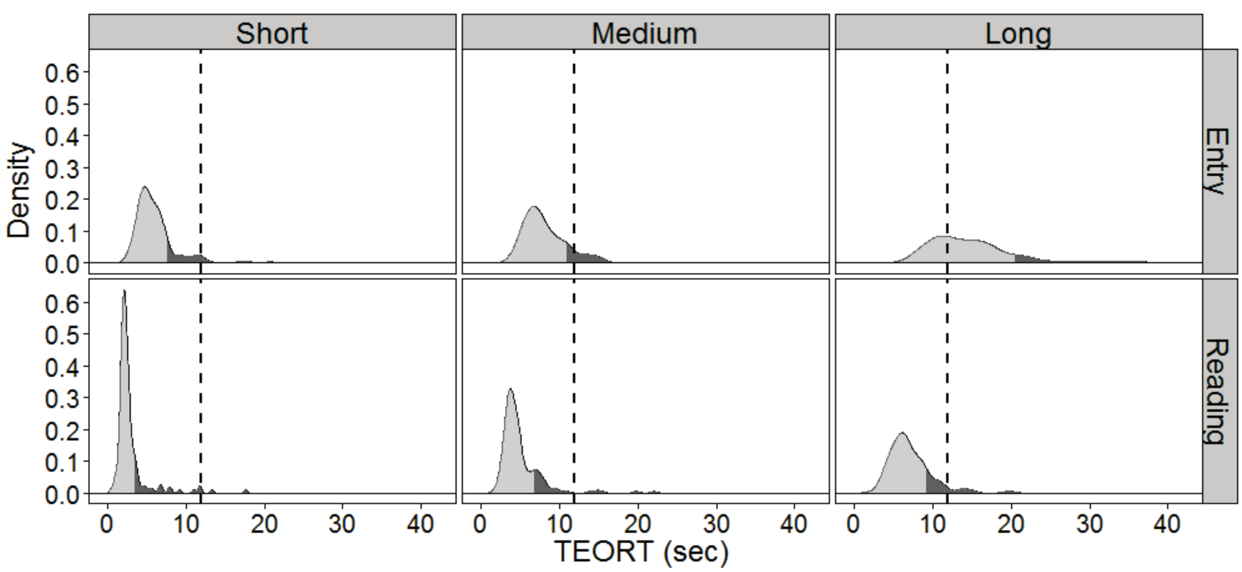
The acceptance criterion for the proposed NHTSA Distraction Guidelines states that for 24 out of 28 participants (85%), the sum of individual glance durations should be less than or equal to 12 seconds.

There were three replications for each test condition and participant. Hence, the mean TEORT of the three replications is used to evaluate whether a participant adhere to the 12-second criterion for any test condition. More specifically, a test condition is considered in conformance if at least 24 out of 28 participants had a mean TEORT ≤ 12 seconds.

3.1.5.1 Descriptive Analysis

Figure 17 shows the density of TEORT for individual tasks, separated by text type and length (There were $N = 168$ data points for each condition). The black vertical dotted lines show the 12-second limit. The light gray area shows the proportion of the distribution that falls below the 85th percentile and the dark gray area shows the upper 15 percent of the distribution. Clearly, the 12-second line falls within the upper 15 percent for all conditions except for long text entry. That is, for all other test conditions shown in the graph except long text entry, no more than 15 percent of tasks exceeded the 12-second limit. However, for long text entry, more than 50 percent of tasks did not conform.

Figure 17 also shows that the distribution mean and variance increase with text length. When examined within each text length, the mean and variance is smaller for text reading than for text entry. The long text entry condition has the largest mean and variance, as well as the longest TEORT. The short text reading condition has the smallest mean and variance.



Note: light gray shows proportion that falls below the 85th percentile; dark gray shows upper 15 percent

Figure 17. Density plot of TEORT by task type and text length at the task level.

When examined at the participant level, some individuals had much longer TEORT values than others. Figure 18 shows the TEORT values for the 12 test conditions, separated by age group. The error bars show the inter-subject standard deviations on the TEORT, and solid dots show the mean TEORT value for each participant. The graphs show that the TEORT value is smaller for text reading than text entry under each sub-test condition. There is a clear increasing trend of TEORT from short to long length, and the trend is larger for text entry than for text reading. Additionally, for text entry, this increase in TEORT is more dramatic from medium to long length than from short to medium length.

The long text entry condition has the largest mean and inter-subject standard deviation for TEORT for all age groups. In fact, most participants had TEORT values above the 12 seconds limit (gray dotted line) under long text entry conditions. Interestingly, although the older participants had the smallest inter-subject standard deviation of MGD for long text entry, they actually had the largest standard deviation for TEORT (Figure 15). That is, although older participants had shorter individual glances to maintain safe driving, some of them still needed longer total glance time to complete the tasks. For text reading, the TEORT values for most participants fell within the 12-second limit. Ambient text did not have as strong an effect as task type or text length; however, it does result in an increase of TEORT for text reading conditions.

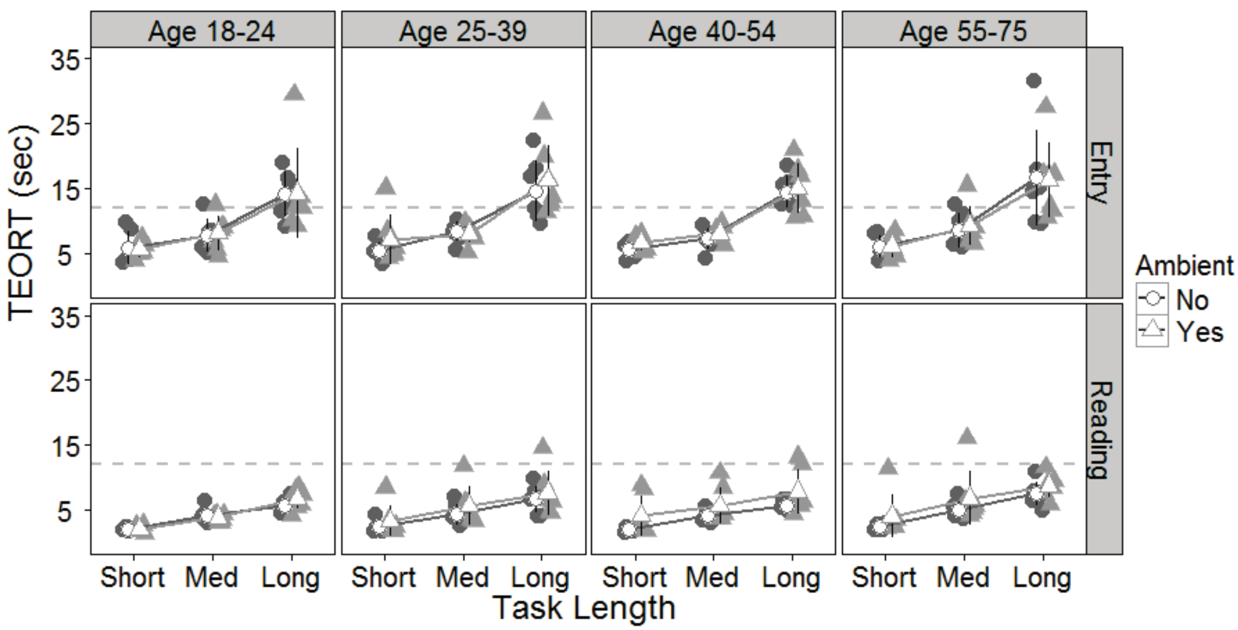


Figure 18. TEORT by age group for each task condition (at the participant level).

Table 12 shows the overall mean of TEORT, 95 percent CI of mean, and inter-subject standard deviations, as well as the number of participants exceeded the 12-second threshold of criterion 3. For text entry, the short and medium length conditions have much smaller overall mean TEORT values. The overall mean of TEORT in long text entry condition is 15.42 seconds (95% CI = [14.25, 16.6]) with ambient text and 14.94 seconds (95% CI = [13.93, 15.95]) without ambient text. However, the inter-subject standard deviations associated with these two means are the largest (*SD* = 5.36s and 4.67s, respectively), which indicate larger individual differences as the text entry length increases. Over 70 percent of participants exceeded the 12-second limit in the long text entry condition regardless of the presence of ambient text, and therefore, the long text entry did not conform to the acceptance criterion. All the other text entry conditions conformed to the acceptance criterion because no more than two participants (7.1%) had a TEORT of more than 12 seconds.

For text reading tasks, all test conditions conformed to acceptance criterion 3, with three or less participants exceeding the threshold. Long text reading with ambient text had the highest number of participants not conforming (*N* = 3) as well as the largest mean of TEORT (*M* = 7.8 seconds). Additionally, all participants were in conformance in text reading without ambient text condition, even in the long text condition.

The Table 12 also shows the number of tasks that exceeded the 12-second limit. More than 60 percent of the individual tasks did not conform to the 12-second limit for the long text entry conditions. Less than 15 percent of tasks did not conform in all other text entry conditions. For text reading, all tasks conformed to the 12-second criterion for the short and medium length without ambient text condition, and only one task did not conform for long length without ambient text. The evaluations on both levels produce the same conclusions for TEORT. However, the nonconformance percentages are slightly different between the participants and task level, which demonstrates the influence of repeated measurements and data aggregation.

Table 12. Total Eyes-Off-Road Time

Task Type	**Length**	**Ambient**	**Mean** (s)	**± 95% CI of mean** (s)	**Inter-subj** *SD* (s)	**No. subj out of 28 (NC)** (and %)	**No. tasks out of 84 (NC)** (and %)
Entry	Short	Yes	6.34	[5.69, 6.99]	2.17	1 (3.6)	4 (4.8)
		No	5.73	[5.27, 6.20]	1.65	0 (0.0)	1 (1.2)
	Med	Yes	8.36	[7.82, 8.90]	2.14	2 (7.1)	8 (9.5)
		No	7.96	[7.43, 8.49]	2.09	2 (7.1)	9 (10.7)
	Long	Yes	15.42	[14.25, 16.6]	5.34	20 (71.4)	57 (67.9)
		No	14.94	[13.93, 15.95]	4.67	21 (75.0)	52 (61.9)
Reading	Short	Yes	3.43	[2.93, 3.94]	2.53	0 (0.0)	2 (2.4)
		No	2.32	[1.93, 2.71]	0.54	0 (0.0)	0 (0.0)
	Med	Yes	5.52	[4.93, 6.11]	2.97	1 (3.6)	5 (6.0)
		No	4.46	[3.95, 4.97]	1.23	0 (0.0)	0 (0.0)
	Long	Yes	7.80	[7.20, 8.40]	2.64	3 (10.7)	7 (8.3)
		No	6.41	[5.91, 6.92]	1.63	0 (0.0)	1 (1.2)

Note: The shaded cells show the test conditions that *did not conform* to the acceptance criterion.
The 95 percent CI of mean is calculated based on intra-subject standard error of mean.

3.1.5.2 ANOVA for Total Eyes-Off-Road Time

A linear mixed model (R 2.12.1) was used to perform the repeated measure ANOVA in TEORT, and the results are shown in Table 13 and Figure 19. The response variable was the log of TEORT to meet the assumptions of ANOVA. The final model was chosen based on the lowest AIC value, and as such, gender and age were not included. Significant differences were observed for all three main effects (Task type, Text length and Ambient) and the interactions effects between task type × text length, and task type × ambient text. Specifically, text entry has significantly larger TEORT value than text reading. For text entry, there is a larger increase in TEORT from medium to long text length than from short to medium text length. The TEORT is greater when ambient text is present and this difference is larger for text reading than for text entry.

Table 13. TEORT ANOVA Results

	Num df	Den df	F-value	p-value
(Intercept)	1	969	2378.47	<.0001
Task Type	**1**	**969**	**254.27**	**<.0001**
Length	**2**	**969**	**868.24**	**<.0001**
Ambient	**1**	**969**	**9.23**	**0.002**
Task Type*Length	**2**	**969**	**21.70**	**<.0001**
Task Type*Ambient	**1**	**969**	**11.86**	**0.001**
Length*Ambient	2	969	1.02	0.36
Task Type*Length*Ambient	2	969	0.12	0.89

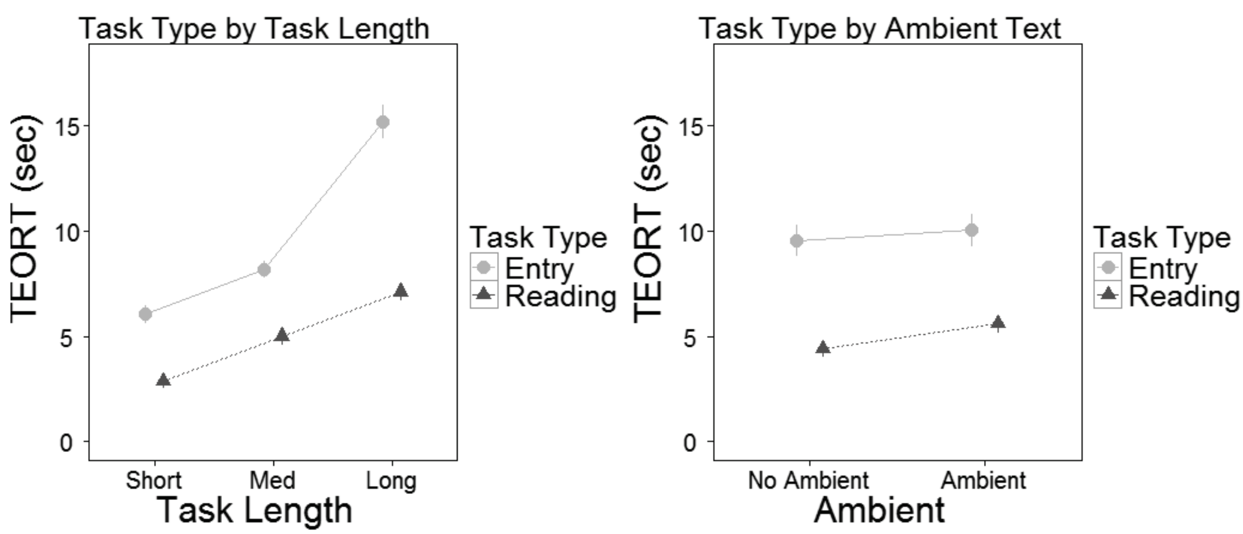


Figure 19. Interaction plots for TEORT.

Note: the error bars show the 95 percent CI for mean based on intra-subject standard error.

3.2 OCCLUSION STUDY

3.2.1 Demographics

The mean age of the participants in the four age groups were 21.3 (*SD* = 0.8, range: 21-23), 30.4 (*SD* = 4.0, range: 26-37), 46.6 (*SD* = 3.6, range: 41-51), and 61.9 (*SD* = 35.6, range: 57-72) years, respectively. All 28 participants had at least high school diplomas, with 9 people (32.1%) having had some college studies, 12 (42.9%) having 4-year college degrees, and 4 (14.3%) having master's degrees or higher.

The mean age that participants started driving was 15.27 years old (*SD* = 0.95 years). The mean age of their first driver's license was 16.23 (*SD* = 0.57 years). None of the participants had a moving violation in the past year; however, four participants had one violation and two participants had two violations in the last three years. Three participants had one vehicular crash (two at fault) and one participant had four vehicular crashes (one at fault) in the last three years. Of these crashes, three had occurred during the past year (none were at fault). The majority of participants (71.4%) reported that they drive at least once daily, 25 percent reported that they drive at least one weekly, and only one person reported to drive less that once weekly.

3.2.2 Task Duration for Occlusion Trials

Figure 20 shows the distribution of the total task duration by task type and text length. A reference line at 15 seconds is included to show the maximum acceptable task duration (this is set by the authors; the proposed NHTSA Distraction Guidelines do not contain a maximum acceptable task duration). In general, entering text took longer to perform than reading text. For text entry, even short words could take more than 15 seconds to complete. The long words shifted the entire distribution toward the right, such that a large portion exceeded the 15-second limit line. For reading tasks, the probability distribution of task duration for short phrases complies with the 15-second limit. For medium-length phrases, there are cases that exceed the 15-second limit (proportion of the distribution to the right of the reference line) and for long phrases, approximately half of the task duration distribution is to the right of the reference line, indicating a large number of cases exceeding the 15-second criterion.

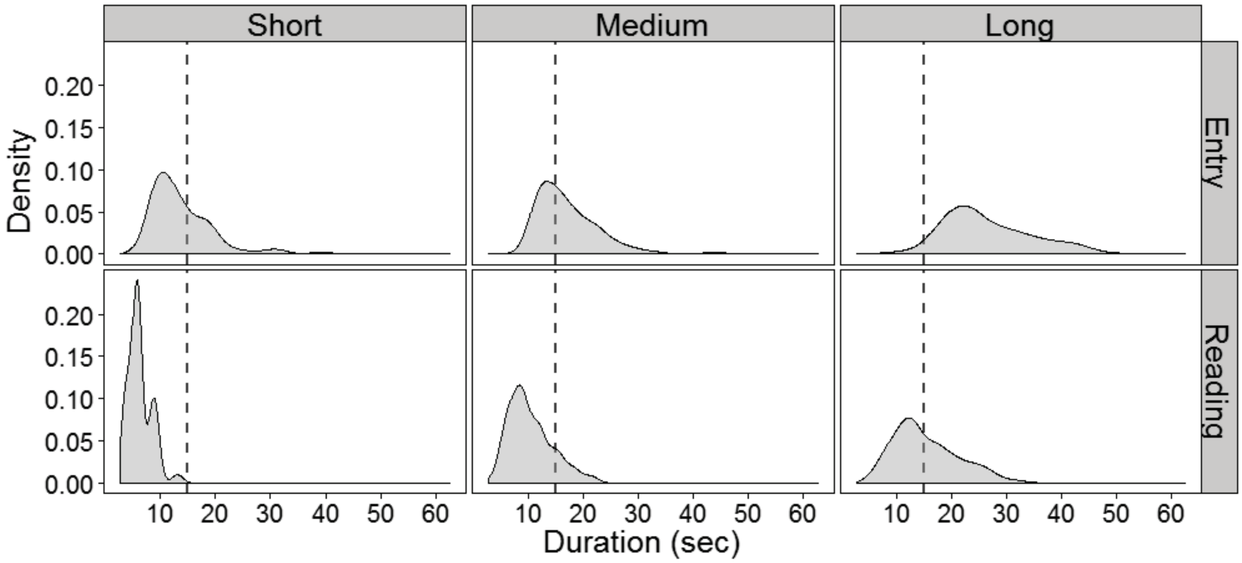


Figure 20. Density plot of task duration in the occlusion condition by task type and length (at the task level).

At the participant level, the mean task duration for the three replications of each test condition is taken as the performance of that participant on that test condition. Figure 21 shows the mean and the inter-subject standard deviation of task duration for each age group, for each task type, text length, and for tasks completed in the presence or absence of ambient text. The 15-second reference line is added to make the task conditions that result in unacceptable task durations distinguishable. For reading tasks, both short and medium-length phrases were read in less than 15 seconds by the majority of participants, whereas for long phrases, and especially for those 25 and older, the means and error bars indicate that there are many cases of 15-second duration violations. For text entry tasks, only the short words have an acceptable profile for the youngest group; however, for the three other age groups, the durations exceed the 15-second limit. For medium phrases, task durations are just above the 15-second margin, with an upward trend when moving from the youngest group to the oldest group. The long reading phrases are particularly problematic, because there is very few cases of task completions in less than 15 seconds, especially for the oldest group whose average task duration is roughly twice as long as the limit.

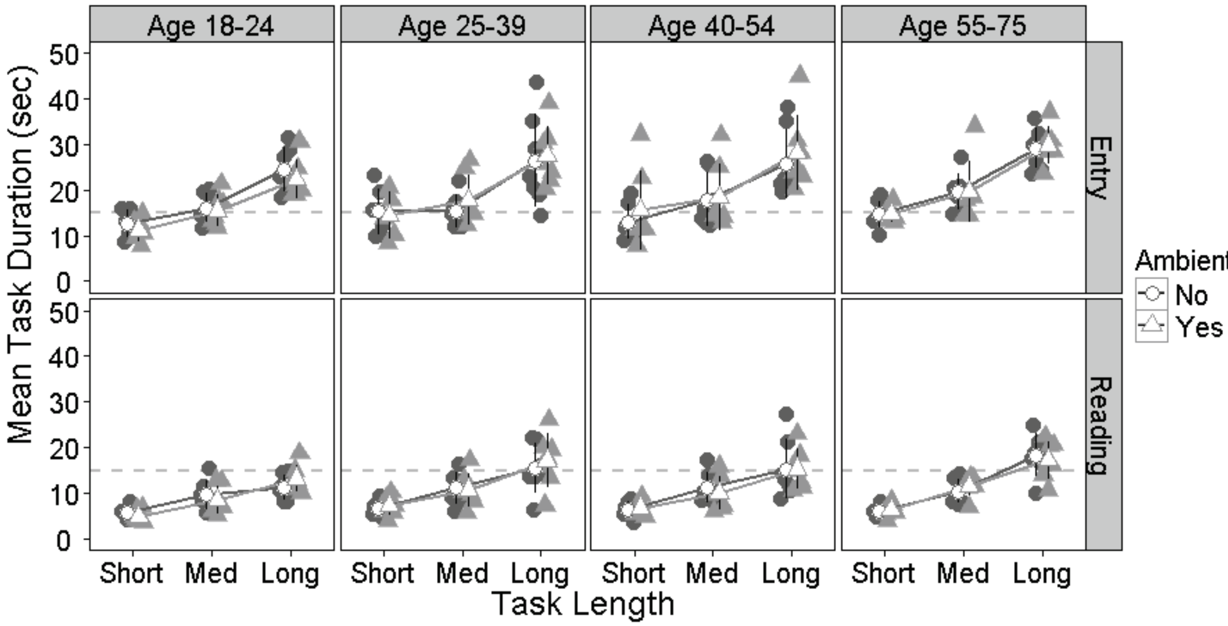


Figure 21. Mean task durations for the occlusion trials by age and task type (at the participant level).

Table 14 provides the summary statistics of task duration times for the occlusion trials, including the number of participants (at the participant and task level) that took longer than 15 seconds to complete the tasks. The mean task duration increased with the text length for both the text entry and text reading conditions. There did not appear to be any difference with ambient text when after controlling for text length and task type.

Text entry tasks appeared to take longer to complete and had larger standard deviations when compared to the text reading tasks. The mean task duration exceeded 15 seconds for the medium and long text length in the task entry condition, and for the long text length in the task reading condition. The longest mean task duration was observed in the long text entry tasks with ambient text (*M* = 26.94s) with all participants in this condition taking longer than 15 seconds to complete the task. When examined at the task level, 98.8 percent (83 out of 84 tasks) in this condition took longer than 15 seconds to complete. The shortest mean task duration (and smallest inter-subject variations) was observed in the short text reading condition without ambient text (*M* = 6.13s, inter-subject *SD* = 1.50s).

Table 14. Task Duration for Occlusion Trials

Task Type	Length	Ambient	Mean (s)	± 95% CI of mean (s)	Inter-subj *SD* (s)	Task duration ≥ 15s: Participants out of 28 (and %)	Task duration ≥ 15s: Tasks out of 84 (and %)
Entry	Short	Yes	13.95	[12.91, 15.00]	5.23	8 (28.6)	25 (29.8)
		No	13.87	[12.86, 14.87]	3.73	10 (35.7)	28 (33.3)
	Med	Yes	17.79	[16.38, 19.20]	5.79	16 (57.1)	53 (63.1)
		No	17.17	[16.15, 18.20]	4.42	16 (57.1)	48 (57.1)
	Long	Yes	26.94	[25.57, 28.30]	6.30	28 (100)	83 (98.8)
		No	26.44	[24.99, 27.88]	6.99	27 (96.4)	81 (96.4)
Reading	Short	Yes	6.30	[5.65, 6.94]	1.79	0 (0.0)	0 (0.0)
		No	6.13	[5.45, 6.82]	1.50	0 (0.0)	0 (0.0)
	Med	Yes	10.13	[9.49, 10.76]	3.27	2 (7.1)	12 (14.3)
		No	10.60	[9.87, 11.33]	3.05	3 (10.7)	13 (15.5)
	Long	Yes	15.70	[14.76, 16.64]	4.53	15 (53.6)	40 (47.6)
		No	15.03	[14.10, 15.95]	5.44	11 (39.3)	33 (39.3)

Note: The 95 percent CI of mean is calculated based on intra-subject standard error of mean.

3.2.3 Criteria 4 and 5: Total Shutter Open Time

This section reports on the results of the TSOT for each task, or the sum of durations of time that the vision is not occluded. The acceptance criteria for TSOT are defined at two levels; 9 and 12 seconds (Criteria 4 and 5). The TSOT for each participant for each task condition is calculated as:

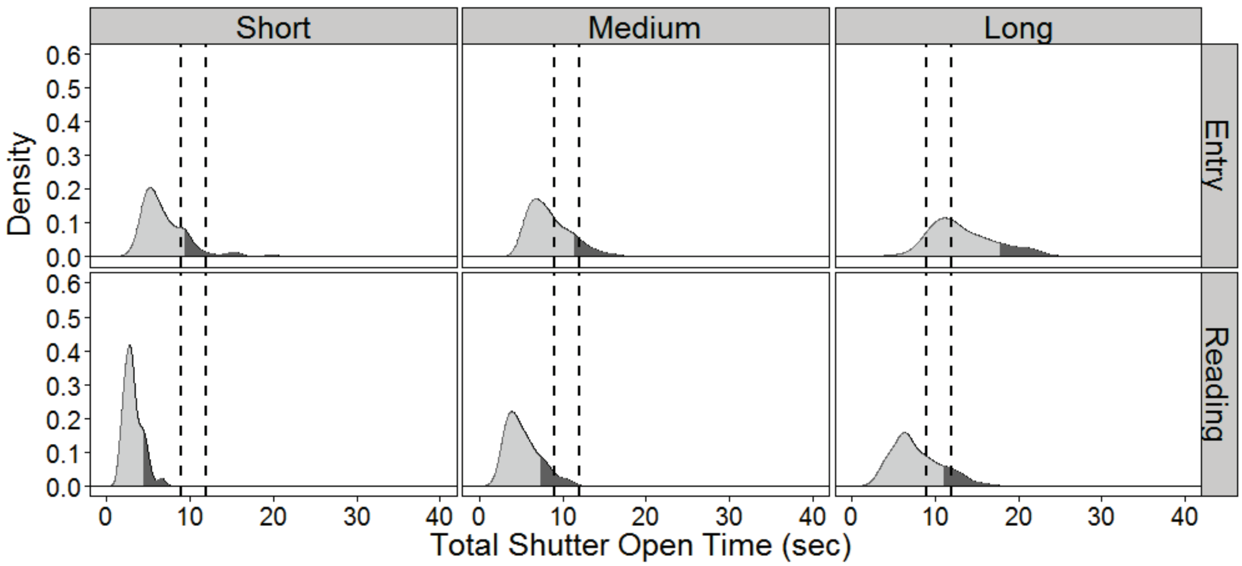
$$TSOT = \frac{\sum_{i=1}^{3} TSOT \text{ in } task_i}{3}$$

If the average TSOT over the three replications of a task condition is less than or equal to 9 (or 12) seconds, the participant's performance is considered to be in conformance with criteria 4 or 5. Otherwise, the participant is not in conformance with criteria 4 or 5. In order for a test condition to be in conformance with 4 or 5, at least 85 percent of the participants should be in conformance. More specifically, for 24 out of 28 participants, the TSOT should be less than or equal to 9 (or 12) seconds.

3.2.3.1 Descriptive Statistics

Figure 22 shows the distribution of TSOTs for each task type and task length. Two reference lines are added to the graph at 9 and 12 seconds to make comparisons against the criteria easier. The light gray area shows the proportion of the distribution that falls below the 85th percentile and the dark gray area shows the upper 15 percent of the distribution. For a condition to conform to the 9 or 12-second criterion, the 85th percentile of the TSOT distribution (in light gray) for that condition should fall to the left of the 9- or 12-second reference line, respectively. For example, short entry tasks conform to the 12-second criterion, but did not conform to the 9-second criterion.

The difference in the distribution of TSOT between short and long tasks (for both reading and entry tasks) is particularly interesting, because not only does the mean task duration increase when moving from short to long tasks, but the variation also increases substantially. In other words, for short tasks (and especially in case of short reading tasks) the performance is relatively similar among the participants whereas for long tasks (particularly for long entry tasks), the participants show considerable variation in the time it takes them to complete the tasks.



Note: light gray shows proportion that falls below the 85th percentile; dark gray shows upper 15 percent.

Figure 22. Density plot of total shutter open time by task type and length (at the task level).

At the participant level, the mean TSOT for the three replications in each task condition is averaged and defined as the performance of that participant on that task condition. Figure 23 plots the mean and inter-subject standard deviation of TSOT for each age group, separated by task type and presence or absence of ambient text. As expected based on the trends, short reading tasks were all completed in less than 9 seconds by all age groups. Medium-length reading tasks were also completed in less than 9 seconds (based on average TSOT for each participant). However, for long reading tasks, one participant 25 or older had an average TSOT greater than 12 seconds when ambient text was presented and two participants had an average TSOT greater than 12 seconds when ambient text was absent. For 9 participants, the 9-second limit was violated when ambient text was present and for 8 participants, this limit was violated when ambient text was absent. For entry tasks, there were many violations of 9- and 12-second limits for short tasks and even a larger number of violations for medium-length tasks. For long tasks, almost all participants did not conform to the 9-second limit and a many did not conform to the 12-second limit.

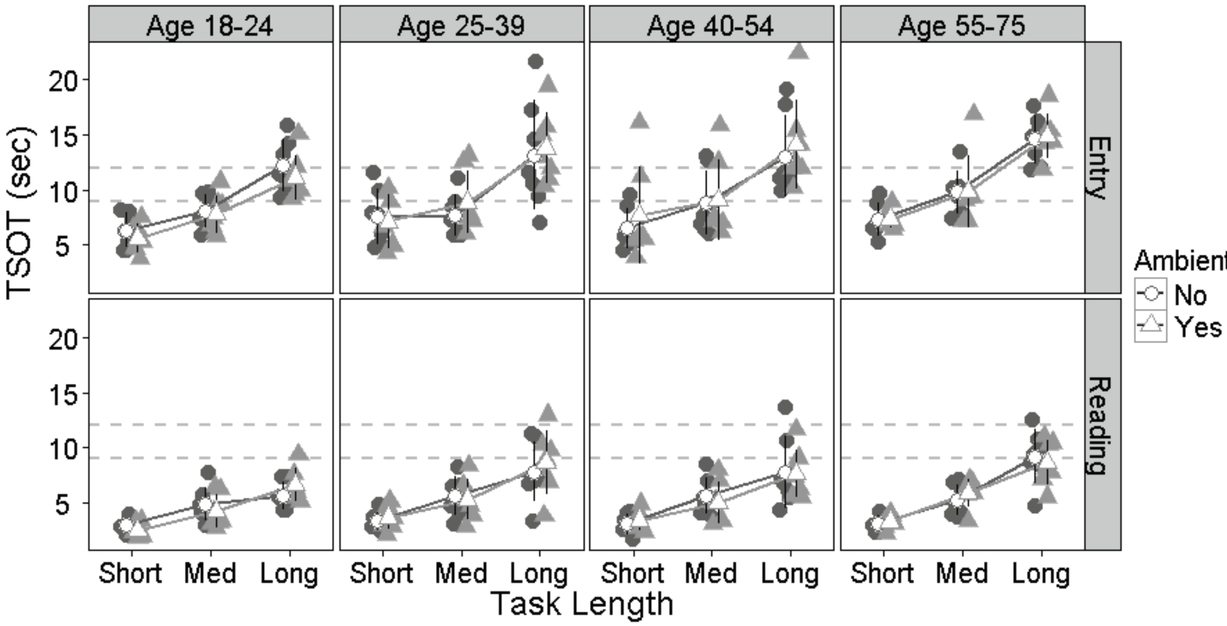


Figure 23. TSOT by age group for each task condition (at the participant level).

Table 15 provides summary statistics for the participants' TSOTs across task types, length, and ambient text. The general trend within both text entry and text reading is that the mean inter-subject *SD* increases as the word/phrase length increases with the largest increase observed for text entry tasks when going from medium to long tasks. There does not appear to be a meaningful difference with and without ambient text (when accounting for task type and length) in terms of mean and inter-subject *SD*.

All participants were in conformance with criteria 4 and 5 for short and medium reading tasks. For the medium reading task, it is noted that two (2.4%) tasks with ambient text and six (7.1%) tasks without ambient text had TSOTs more than 9 seconds (although less than 12 seconds). None of the test conditions associated with the long text reading task was in conformance with criterion 4, but they were in conformance with criterion 5.

For text entry, all short text entry conditions were in conformance with criterion 5. For criterion 4, the short text entry, when examined at the task level, was not in conformance. None of the medium text entry conditions were in conformance with criterion 4, but almost all were in conformance with criterion 5. At the participant level, the medium text length with ambient text was not in conformance. Regardless of analysis at the participant or task level, no long text entry tasks were in conformance with criterion 4 or 5.

Table 15. Number of Participants and Tasks that are not compliant (NC) based on criteria 4 and 5 for each test condition.

Task Type	Length	Ambient	Mean TSOT (s)	± 95% CI	Inter-subj *SD* (s)	9-second criterion: Participants NC out of 28 (and %)	9-second criterion: Tasks NC out of 84 (and %)	12-second criterion: Participants NC out of 28 (and %)	12-second criterion: Tasks NC out of 84 (and %)
Entry	Short	Yes	6.95	[6.43 , 7.47]	2.62	4 (14.3)	18 (21.4)	1 (3.6)	4 (4.8)
		No	6.96	[6.46 , 7.45]	1.87	4 (14.3)	17 (20.2)	0 (0.0)	5 (6.0)
	Med	Yes	8.9	[8.20 , 9.61]	2.89	10 (35.7)	30 (35.7)	5 (17.9)	10 (11.9)
		No	8.59	[8.09 , 9.10]	2.2	10 (35.7)	30 (35.7)	3 (10.7)	9 (10.7)
	Long	Yes	13.49	[12.81 , 14.17]	3.13	28 (100)	77 (91.7)	16 (57.1)	51 (60.7)
		No	13.27	[12.55 , 13.99]	3.44	27 (96.4)	75 (89.3)	14 (50.0)	47 (56.0)
Reading	Short	Yes	3.19	[2.87 , 3.51]	0.9	0 (0)	0 (0.0)	0 (0.0)	0 (0.0)
		No	3.09	[2.75 , 3.43]	0.75	0 (0)	0 (0.0)	0 (0.0)	0 (0.0)
	Med	Yes	5.09	[4.77 , 5.41]	1.66	0 (0)	2 (2.4)	0 (0.0)	0 (0.0)
		No	5.3	[4.93 , 5.67]	1.54	0 (0)	6 (7.1)	0 (0.0)	0 (0.0)
	Long	Yes	7.88	[7.42 , 8.35]	2.23	9 (32.1)	25 (29.8)	1 (3.6)	8 (9.5)
		No	7.6	[7.14 , 8.06]	2.71	8 (28.6)	27 (32.1)	2 (7.1)	9 (10.7)

Notes: NC: not compliant, Shaded cells show test conditions that *conformed to* the acceptance criterion. The 95 percent CI of mean is calculated based on intra-subject standard error of mean.

Table 16 presents TSOT percentiles at the participant level, i.e., based on mean TSOT computed over the three replications of each task condition completed by the participant. The 50th percentile or median values are close to the mean values (see Table 15). The 85th percentiles signify the tail of the TSOT distribution, i.e., only 15 percent of the distribution has TSOT values larger than them.

Table 16. Total Shutter Open Time Percentiles for Different Task Conditions

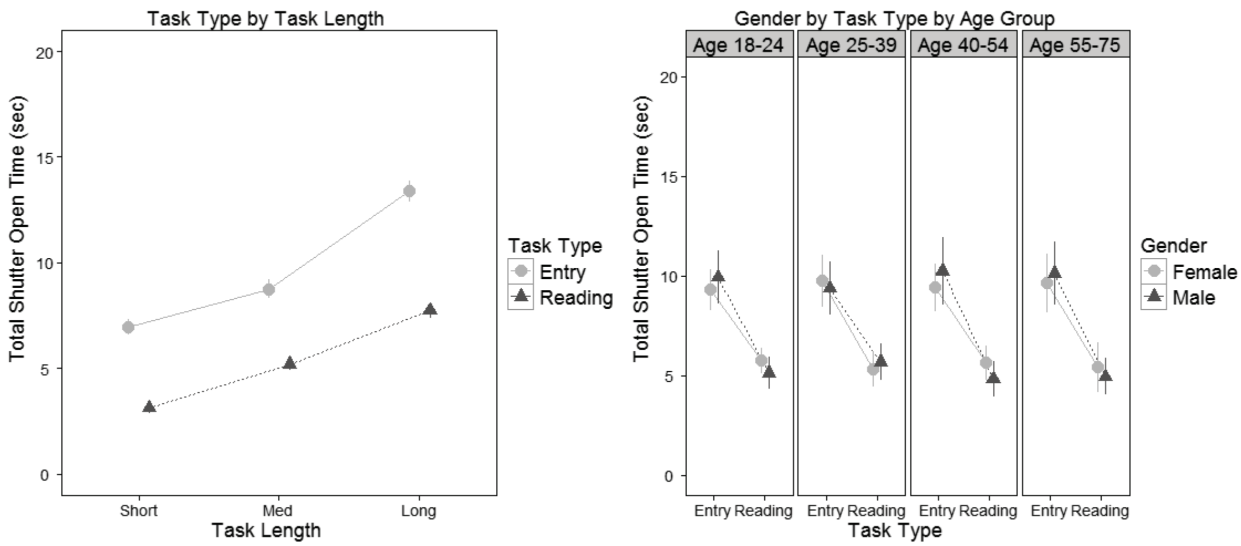
Task Type	Length	Ambient	Percentiles: 25th	50th (Median)	75th	85th
Entry	Short	Yes	5.27	6.52	7.75	8.97
		No	5.36	6.55	8.14	8.77
	Med	Yes	7.21	7.66	9.41	12.48
		No	7.01	8.24	9.74	10.38
	Long	Yes	10.98	12.87	15.19	15.45
		No	10.96	12.01	15.90	17.25
Reading	Short	Yes	2.36	3.17	3.73	4.11
		No	2.60	2.85	3.62	4.01
	Med	Yes	3.38	5.24	6.30	6.83
		No	4.13	4.86	6.51	6.97
	Long	Yes	6.18	7.72	9.50	10.26
		No	5.50	7.15	9.56	10.76

3.2.3.2 ANOVA for Total Shutter Open Time

Linear mixed model (R 2.12.1) was used to perform the repeated measures ANOVA on TSOT. The response variable TSOT was log transformed to meet the ANOVA assumptions. The ANOVA results are shown in Table 17 and Figure 24. The within-subject factors task type and task length have significant main and interaction effects (see the interaction plot on the left of Figure 24). Specifically, text entry has a larger TSOT than text reading. The main effect of ambient text on TSOT was not significant. However, there was a significant three-way interaction indicating that there were differences in TSOT based on the combination of age, gender, and type of distracting task. The biggest differences observed were that middle-aged (40 to 54 years old) and older (55 to 75) males had the longest TSOT for text entry but these same age/gender groups also had the shortest TSOT for text reading.

Table 17. Total Shutter Open Time ANOVA Results

	Num df	Den df	F-value	p-value
(Intercept)	1	962	2180.15	<.0001
Task Type	**1**	**962**	**722.91**	**<.0001**
Gender	1	20	1.16	0.29
Age Group	3	20	1.55	0.23
Length	**2**	**962**	**817.04**	**<.0001**
Ambient	1	962	0.5	0.48
Gender: Age Group	3	20	0.15	0.93
Length*Ambient	2	962	1.01	0.37
Task Type*Gender	1	962	0.36	0.55
Task Type*Age Group	3	962	1.34	0.26
Task Type*Length	**2**	**962**	**23.66**	**<.0001**
Task Type*Ambient	1	962	0.01	0.93
Task type: Gender*Age Group	**3**	**962**	**3.47**	**0.02**
Task Type*Length*Ambient	2	962	1.26	0.28



Note: the error bars show the 95 percent CI for mean based on intra-subject standard error.

Figure 24. Interaction plots for total shutter open time.

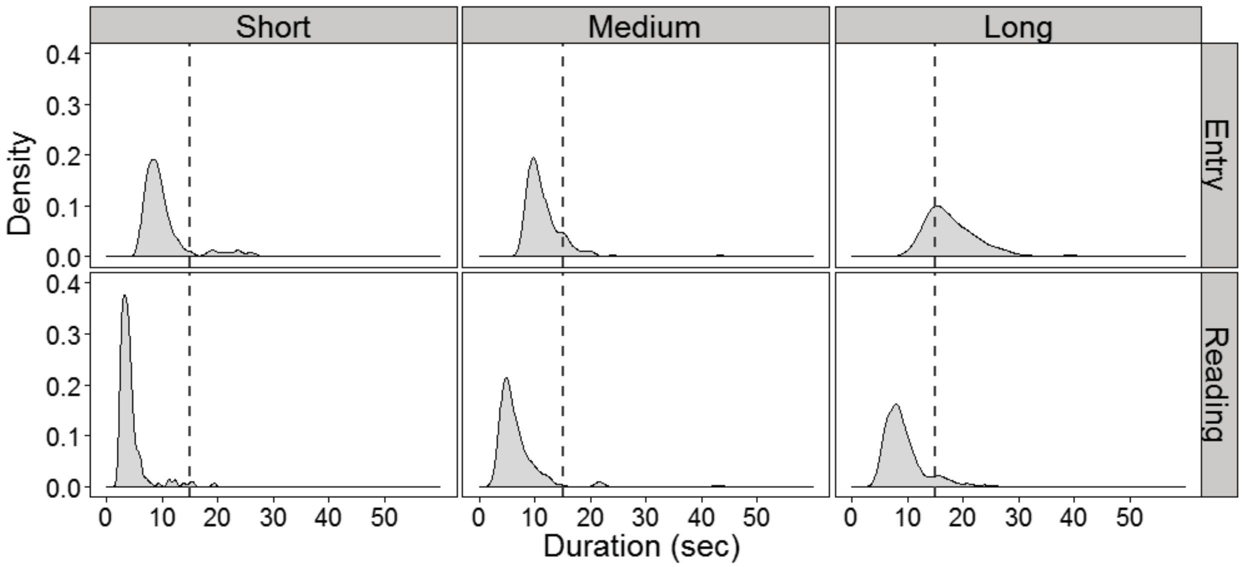
3.3 COMPARING DRIVING SIMULATOR AND OCCLUSION STUDY

The driving simulator study and the visual occlusion study were conducted at two different institutions, University of Washington and University of Wisconsin-Madison. Both studies included static trials. In the simulator study, the static trials involved completing text entry and text reading tasks when the simulator vehicle was in the parked position (i.e., without driving). In the occlusion study, static trials involved completing text entry and reading tasks without wearing the occlusion goggles. The order of simulator (or occlusion) trials and static trials were randomized, such that for half of the participants at each location the static trials came before the simulator (occlusion) trials and for the other half came after the simulator (occlusion) trials.

In designing the experimental protocol at the two institutions, the research team made the two settings as similar as possible. Most important, the location of the touchscreen relative to the steering wheel and driver seat was matched at the two study locations. In addition, the exact same touchscreen model and experimental program was used and all study procedures were coordinated so that the maximum level of consistency is achieved. However, it would still be worthwhile to compare the performance of participants at the two locations (28 at each location) in the static trials to see if any substantial difference is observed.

Figure 25 shows the distribution of total task duration for static trials, separated by task type and task length for the simulator study (Figure 25a) and the occlusion study (Figure 25b). A reference point is drawn at 15 seconds to facilitate comparisons. Note that the task duration distributions are very similar across the locations, although minor differences exist. For example, there is more variation in the task durations of short and medium reading tasks for participants in the static condition in the Seattle, WA study compared to the Madison, WI area. There are other small differences that can be found by comparing the distributions shown in the two panels of Figure 25; however, the general trends in task duration are the same in the two locations.

(a) University of Washington: Simulator study



(b) University of Wisconsin-Madison: Occlusion study

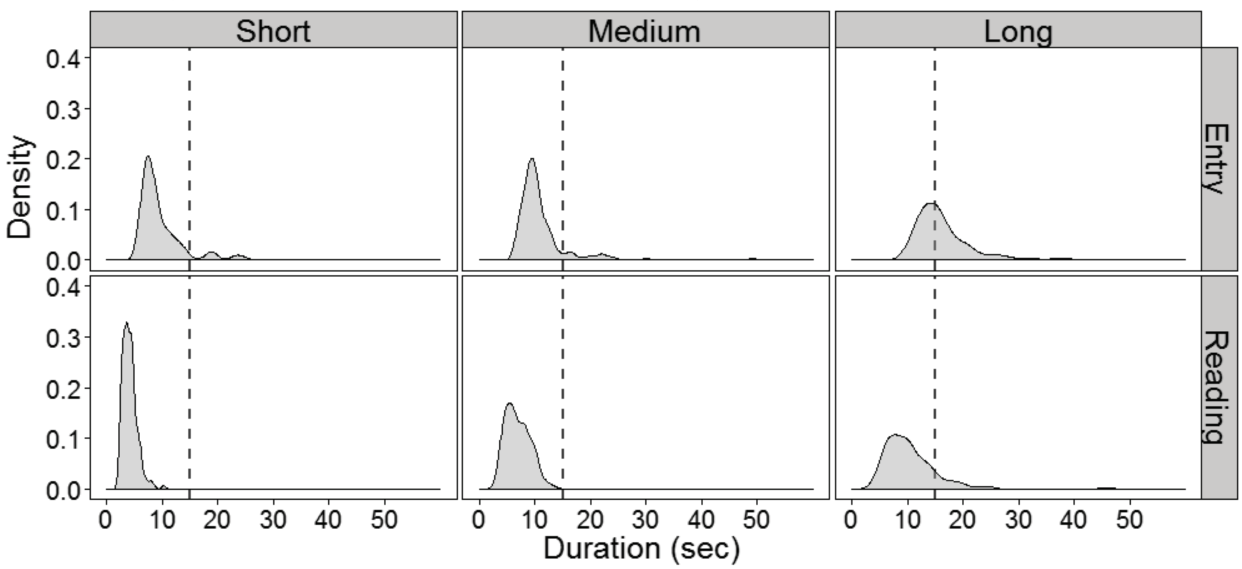
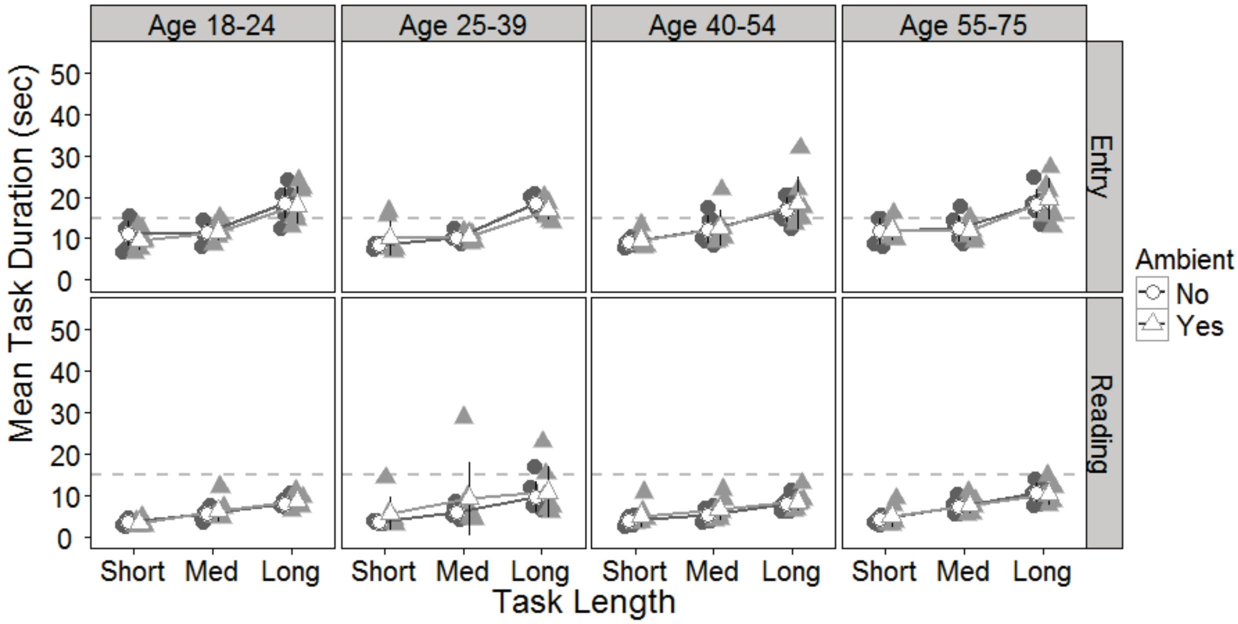


Figure 25. Density plot of task duration in static trials by task type and length (at the task level).

Figure 26 plots the means and inter-subject standard deviations, separated based on age groups and task types and drawn by task length and ambient text. The horizontal line at 15 seconds serves to facilitate comparison between different age groups and task conditions, across the two study locations. Here again, the general patterns are similar and consistent with expectations, although not identical.

(a) University of Washington: Simulator study



(b) University of Wisconsin-Madison: Occlusion study

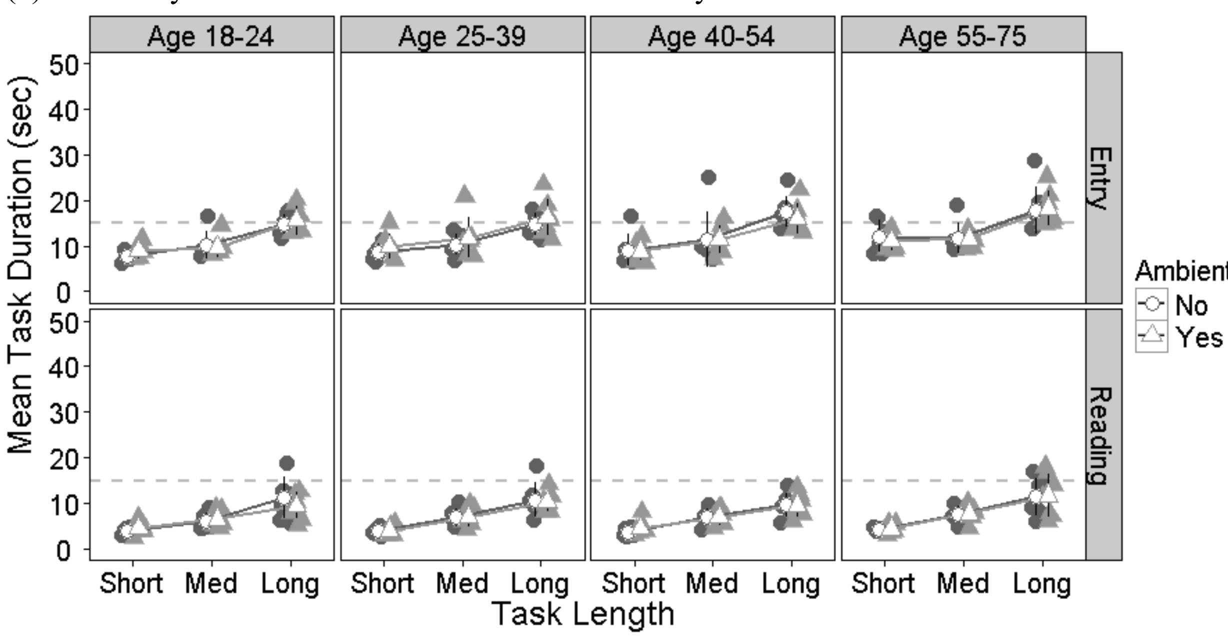


Figure 26. Mean task durations in static trials by age group, for each task condition (at the participant level).

Table 18 and Table 19 summarize task durations for the different conditions in the driving simulator and the occlusion goggles, respectively. Although the general performance is similar between the two protocols, there are some noted differences. For example, the long text entry tasks took on average two seconds longer to complete in the simulator study compared to occlusion study, for each of with ambient text and without ambient text conditions. There are also differences in variation of task duration. For example, the total standard deviation of the duration of medium reading tasks with ambient text is 5.42 seconds in the simulator study and 2.17 seconds in the occlusion study, although the mean durations for this condition are very similar (7.43 sec and 7.01 sec, respectively).

Table 18. Driving Simulator: Total Task Durations in Static Trials

	Length	Ambient	Mean(s)	± 95% CI of mean (s)	Inter-subj *SD* (s)	Task duration ≥ 15s	
						No. subj (%, $N = 28$)	No. tasks (%, $N = 84$)
Entry	Short	Yes	10.29	[9.39 , 11.18]	2.88	3 (10.7)	10 (11.9)
		No	10.12	[9.24 , 11.01]	2.51	1 (3.6)	6 (7.1)
	Med	Yes	11.65	[10.77 , 12.53]	2.71	3 (10.7)	9 (10.7)
		No	11.56	[10.87 , 12.25]	2.64	2 (7.1)	12 (14.3)
	Long	Yes	18.23	[17.01 , 19.44]	4.64	20 (71.4)	56 (66.7)
		No	18.12	[17.18 , 19.07]	3.01	24 (85.7)	62 (73.8)
Reading	Short	Yes	4.71	[4.05 , 5.36]	2.73	0 (0.0)	3 (3.6)
		No	3.78	[3.37 , 4.20]	0.77	0 (0.0)	0 (0.0)
	Med	Yes	7.43	[6.36 , 8.51]	4.79	1 (3.6)	4 (4.8)
		No	6.03	[5.49 , 6.58]	1.63	0 (0.0)	0 (0.0)
	Long	Yes	9.55	[8.84 , 10.26]	3.55	2 (7.1)	9 (10.7)
		No	9.07	[8.36 , 9.79]	2.59	1 (3.6)	7 (8.3)

Note: The 95 percent CI of mean is calculated based on intra-subject standard error of mean.

Table 19. Occlusion Goggles: Total Task Durations in Static Trials

	Length	Ambient	Mean(s)	± 95% CI of mean (s)	Inter-subj *SD* (s)	Task duration ≥ 15s	
						No. subj (%;$N = 28$)	No. tasks (%;$N = 84$)
Entry	Short	Yes	9.66	[8.79 , 10.52]	2.27	1 (3.6)	8 (9.5)
		No	9.40	[8.69 , 10.10]	2.99	3 (10.7)	5 (6.0)
	Med	Yes	11.07	[10.26 , 11.88]	2.99	2 (7.1)	10 (11.9)
		No	10.87	[9.78 , 11.95]	3.80	3 (10.7)	7 (8.3)
	Long	Yes	16.49	[15.40 , 17.58]	3.58	16 (57.1)	44 (52.4)
		No	16.16	[15.21 , 17.12]	3.63	16 (57.1)	43 (51.2)
Reading	Short	Yes	4.34	[3.98 , 4.70]	1.23	0 (0.0)	0 (0.0)
		No	3.88	[3.50 , 4.26]	0.68	0 (0.0)	0 (0.0)
	Med	Yes	7.01	[6.59 , 7.43]	1.78	0 (0.0)	0 (0.0)
		No	6.84	[6.41 , 7.27]	1.83	0 (0.0)	0 (0.0)
	Long	Yes	10.22	[9.49 , 10.96]	3.20	2 (7.1)	8 (9.5)
		No	10.67	[9.46 , 11.88]	3.67	4 (14.3)	11 (13.1)

Note: The 95 percent CI of mean is calculated based on intra-subject standard error of mean.

In summary, although making comparisons between the outcomes of experiments at the two study locations was not an aim of this study, this section briefly described the results of the static trials, which followed the same protocol at the two locations. This comparison showed major similarities but also revealed minor disparities. However, given the relatively small sample size, it is difficult to conclude whether the differences noted in this study exist for larger populations.

4 CONCLUSIONS AND DISCUSSION

The purpose of this study was to evaluate the limits of drivers' ability to perform text entry and reading tasks while driving, and also consider other factors that affect distraction associated with text entry and reading. The goal was to provide information that can be useful for the proposed NHTSA Visual-Manual Driver Distraction Guidelines with respect to the text entry and text reading specifications.

Two studies were conducted with 28 participants each: a driving simulator study and an occlusion study. The driver simulator study showed that task type, task length, and ambient text influenced the mean and total eyes-off-road durations. MGD duration and TEORT were significantly longer for long text string lengths when compared to short text string lengths. Drivers engaged in the long text entry tasks (12 characters) had significantly longer TEORT values when compared to the medium (6 characters) and short (4 characters) text entry tasks. Ambient text increased TEORT values by about 1.0 seconds for text reading and 0.5 seconds for text entry; a much smaller effect than text length and task type. It is important to note that there are many possible types of ambient text, and this current study examined only one type. The effects of other types of ambient text may be different and further studies to investigate would be useful.

Some extremely long off-road glances were observed in the simulator study, which have rarely been observed in previous studies. This may be due to the experimental protocol used (e.g., the specific secondary tasks, the headway distance in the driving scenarios). Ad-hoc observations showed that some participants tend to have longer off-road eye glances because they could enter several characters in one chunk (or in one long glance), and this behavior did not necessarily change when the text entry tasks were shorter. Therefore, shortening the length of the text entry tasks may not actually guarantee that the EOR duration would be shorter (or within a safe zone) for all drivers. However, the longer off-road glances may have increased the number of lane departures.

The results of the occlusion study showed significant differences for task type and task length. More specifically, TSOT was significantly longer for text entry tasks when compared to text reading tasks of comparable length levels (short, medium, and long). TSOT also increased as the length of the reading or entry tasks increased. Similar to the results obtained from the simulator study, TSOT values were longer for the long text entry tasks (did not conform to criterion 4 and 5) compared to short and medium text entry tasks; although even medium entry tasks with ambient text did not conform with either criterion 4 or 5, and medium entry tasks without ambient text did not conform with criterion 4. The short and medium text reading tasks conformed to criteria 4 and 5; but the long text reading tasks did not conform to criterion 4.

Table 20 summarizes the evaluation results for the five acceptance criteria. All text entry tasks, regardless of length and the presence of ambient text, did not conform with criteria 1 and 2. The short and medium text entry tasks conformed to criterion 3, the short text entry tasks conformed to criterion 4 and 5, and the medium entry tasks without ambient text conformed with criterion 5. All text reading tasks, regardless of length and the presence of ambient text, conformed to criteria 2, 3 and 5. Short and medium text reading conformed to criterion 4. They also both conformed to criterion 1 if conformance was based on performing at least 1 of the 3 replications in conformance.

Table 20. Summary of Acceptance Criteria Evaluation

			Simulator Study			Occlusion Study	
	Length	Ambient	**Criterion 1** long EOR <15%	**Criterion 2** Mean EOR<2s	**Criterion 3** Sum EOR≤12s	**Criterion 4** TSOT≤9s	**Criterion 5** TSOT≤12s
Entry	Short	Yes	NC	NC	C	C	C
		No	NC	NC	C	C	C
	Med	Yes	NC	NC	C	NC	NC
		No	NC	NC	C	NC	C
	Long	Yes	NC	NC	NC	NC	NC
		No	NC	NC	NC	NC	NC
Reading	Short	Yes	C	C	C	C	C
		No	C	C	C	C	C
	Med	Yes	C	C	C	C	C
		No	C	C	C	C	C
	Long	Yes	NC	C	C	NC	C
		No	NC	C	C	NC	C

Note: The shaded cells with a "C" indicate the test conditions that *conformed to* the acceptance criterion.
C: Compliant, NC: Not compliant

The effect of ambient text differed between the simulator and occlusion study with respect to TEORT and TSOT. TEORT and TSOT both are measures of time engaged in a task; TEORT measures the time looking off the road in the simulator, while TSOT measures the time vision is not occluded by the goggles. The presence of ambient text increased TEORT values, especially for text reading tasks, but had no impact on TSOT in the occlusion task. This research showed that drivers that entered text in the driving simulator study did not conform to NHTSA criteria 1 and 2, even for words as short as four characters. Text entry and text reading tasks conformed to the other three criteria (3, 4, and 5) for the short and medium text reading task, and for the short text entry task. With respect to the 30-character limitation as described in the proposed NHTSA Visual-Manual Driver Distraction Guidelines, this study showed that the only task that conformed to all five criteria with this character limitation (20 to 40 characters) was the text reading task with no ambient text. Drivers were able to conform to four of the five criteria in the text reading condition for up to an 80-character display. The results also showed that ambient text does not necessarily impact each drive equally for text entry and reading. In summary, the number of criteria that were met is highly dependent on the context of the distracting task (reading or entering text) and the number of characters displayed. The proposed guidelines for adherence to the 30 character limit would limit the amount of information displayed on an in-vehicle information system. However, increasing the number of characters may make it difficult to adhere to all 5 criteria.

A direct comparison of the secondary tasks between the simulator and occlusion protocol showed that the use of a 12-second criterion for the occlusion TSOT (criterion 5) provided task acceptable results that were more consistent with the 12-second TEORT criterion (criterion 3) for driving glances) than for the 9-second TSOT (Criterion 4). Furthermore, this research showed that the mean ratio between TSOT and TEORT for the Text Entry task, across all text length and ambient text conditions, was 1.04. The corresponding mean ratio for the Text Reading task was 1.09.

REFERENCES

Alliance of Automobile Manufacturers. (2006). *Statement of principles, criteria and verification procedures on driver interactions with advanced in-vehicle information and communication systems*. Washington, DC: Author.

Angell, L., Auflick, J., Austria, P. A., Kochhar, D., Tijerina, L., Biever, W., … & S. Kiger. (2006, November). *Driver Workload Metrics Project, Task 2 Final Report*. (Report No. DOT HS 810 635). Washington, DC: National Highway Traffic Safety Administration. Available at www.nhtsa.gov/DOT/NHTSA/NRD/Multimedia/PDFs/Crash%20Avoidance/Driver%20 Distraction/Driver%20Workload%20Metrics%20Final%20Report.pdf

Ascone, D., Lindsey, T., & Varghese, C. (2009, September). *An examination of driver distraction as recorded in NHTSA databases*. (Traffic Safety Facts Research Note. Report No. DOT HS 811 216). Washington, DC: National Highway Traffic Safety Administration. Available at www-nrd.nhtsa.dot.gov/Pubs/811216.pdf

Brumby, D. P., Salvucci, W., Mankowski, D. D., & Howes, A. (2007). A cognitive constraint model of the effects of portable music-player use on driver performance. In *Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting*, 1531-1535, Baltimore, MD, Oct. 1-5, 2007.

Campbell, J., Carney, C., & Kantowitz, B. (1998). *Human Factors Design Guidelines for Advanced Traveler Information Systems (ATIS) and Commercial Vehicle Operations (CVO)* (Report No. FHWA-RD-98-057). McLean, VA: Federal Highway Administration. Available at www.fhwa.dot.gov/publications/research/safety/98057/

Chen, H., & Tsoi, K. (1988). Factors Affecting the Readability of Moving Text on a Computer Display. *Human Factors, 30*(1), 25-33.

Dingus, T. A., Antin, J. A., Hulse, M., & Wierwille, W. W. (1989). Attentional demand requirements of an automobile moving-map navigation system. *Transportation Research Part A, 23*(4), 301-315.

Donmez, B., Boyle, L., & Lee, J. D. (2007a). Accounting for time-dependent covariates in driving simulator studies. *Theoretical Issues in Ergonomics Sciences*, 1-11.

Donmez, B., Boyle, L., & Lee, J. D. (2007b). Safety implications of providing real-time feedback to distracted drivers. *Accident Analysis & Prevention, 39*(3), 581-590.

Donmez, B., Boyle, L., & Lee, J. D. (2008). Mitigating driver distraction with retrospective and concurrent feedback. *Accident Analysis and Prevention, 40*(2), 776-786.

Donmez, B., Boyle, L., & Lee, J. D. (2010). Differences in off-road glances: effects on young drivers' performance. *Journal of Transportation Engineering, 136*(5), 403-410.

Duchnicky, R., & Kolers, P. (1983). Readability of Text Scrolled on Visual Display Terminals as a Function of Window Size. *Human Factors, 25*(6), 683-692.

Dudek, C. L. (1992). *Guidelines on the Selection and Design of Messages for Changeable Message Signs* (Report No. FHWA-TX-92-1232-10). McLean, VA: Federal Highway Administration.

Dudek, C. L. (1997). NCHRP Synthesis 237: Changeable Message Signs. Washington, DC: Transportation Research Board.

Dudek, C. L. (1999). Changeable Message Sign Messages for Work Zones: Time of Day, Days of Week, and Month Dates. *Transportation Research Record, 1692*, 1-8.

Dudek, C. L. (2004). *Changeable Message Sign Operation and Messaging Handbook.* (Report No. FHWA-OP-03-070). Washington, DC: Federal Highway Administration.

Dudek, C. L. (2008). *NCHRP Synthesis 383: Changeable Message Sign Displays during Non-Incident, Non-Roadwork Periods*. Washington, DC: Transportation Research Board.

Dudek, C. L., Schrock, S. D., & Ullman, B. D. (2007). License Plate and Telephone Numbers in Changeable Message Sign ABMER Alert Messages. *Transportation Research Record: Journal of the Transportation Research Board, 2012*, 64-71.

Dudek, C. L., Schrock, S. D., Ullman, G. L., & Chrysler, S. T. (2006). Flashing Message Features on Changeable Message Signs. *Transportation Research Record: Journal of the Transportation Research Board, 1959*, 122-129.

Fletcher, C., & Chrysler, S. T. (1990). Surface forms, textbases and situation models: Recognition memory for three types of textual information. *Discourse Processes, No. 13*, 175-190.

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C. H., Jones, G., Oliver, I., et al. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences, 5*(6), 236-243.

Greenstein, J. S., & Arnaut, L. Y. (1987). Chapter 11.4: Human Factors Aspects of Manual Computer Input Devices. In G. Salvendy (Ed.), *Handbook of Human Factors* (pp. 1450-1489). New York: J. Wiley & Sons, Inc.

Harbluk, J., Burns, P., Go, E., & Morton, A. (2006). The occlusion procedure for assessing in-vehicle telematics: Tests of current vehicle systems. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, 2373-2377, San Francisco, CA, Oct. 16-20, 2006.

Hoffman, J. D., Lee, J. D., & McGehee, D. (2006). *Dynamic display of in-vehicle text messages: The impact of varying line length and scrolling rate.* Paper presented at the 50th Annual Meeting of the Human Factors and Ergonomics Society, San Francisco, CA, Oct. 16-20, 2006.

Horrey, W. J., & Wickens, C. (2007). In-vehicle glance duration: distributions, tails, and model of crash risk. *Transportation Research Record, 2018*(1), 22-28.

Horrey, W. J., & Wickens, C. D. (2004). Driving and side task performance: The effects of display clutter, separation, and modality. *Human Factors, 46*(4), 611-624.

Horrey, W. J., & Wickens, C. D. (2006). Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human Factors, 48*(1), 196-205.

Horrey, W. J., Wickens, C. D., & Consalus, K. P. (2006). Modeling drivers' visual attention allocation while interacting with in-vehicle technologies. *Journal of Experimental Psychology: Applied, 12*(2), 67.

ISO. (2005). Road Vehicles–Ergonomic aspects of transport information and control systems – Occlusion method to assess visual distraction due to the use of in-vehicle information and communication systems. Draft International Standard. ISO/DIS 16673.

ISO. (2007). Road Vehicles -- Ergonomic Aspects of Transport Information and Control Systems -- Occlusion method to assess visual demand due to the use of In-vehicle Systems. International Standard 16673.

Japan Automobile Manufacturers Association, Inc.. (2004). *JAMA Guideline for In-Vehicle Display Systems, Version 3.0*. Tokyo: Author. Available at www.umich.edu/~driving/documents/JAMA_guidelines_v30.pdf

Juola, J., Tiritoglu, A., & Pleunis, J. (1995). Reading text presented on a small display. *Applied Ergonomics, 26*(3), 221-229.

Kamiya, H., Nakamura, Y., & Matsumoto, H. (1994). A Study on Recognition of In-Car Visual Information. In Vehicle Navigation and Information Systems Conference, Proceedings., 1994 , 469-472. Yokahama, Japan, Aug. 31-Sept. 2, 1994.

Kang, J., & Muter, P. (1989). Reading dynamically displayed text. *Behaviour & Information Technology, 8*(1), 33-42.

Klare, G. R. (1974). Assessing readability. *Reading research quarterly*, 62-102.

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsay, D. J. (2006). *The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data*. (Report No. DOT HS 810 594). Washington, DC: National Highway Traffic Safety Administration. Available at www.nhtsa.gov/DOT/NHTSA/NRD/Multimedia/PDFs/Crash%20Avoidance/Driver%20 Distraction/810594.pdf

Labiale, G. (1996). Complexity of in-car visual messages and driver's performance. In A.G. Gale et al. (Ed.), *Vision in Vehicles* (pp. 187-194). Bron Cedex, France: INRETS.

Lamble, D., Laakso, M., & Summala, H. (1999). Detection thresholds in car following situation and peripheral vision: Implications for positioning of visually demanding in-car displays. *Ergonomics, 42*, 807-815.

Lee, J. D., Young, K. L., & Regan, M. A. (2008). Defining driver distgraction. In M. A. Regan, J. D. Lee & K. L. Young (Eds.), *Driver distraction: Theory, effects, and mitigation* (pp. 31-40). Boca Raton, FL: CRC Press, Taylor & Francis Group.

Lerner, N., Singer, J., Robinson, E., Huey, R., & Jenness, J. (2009, July). *Driver Use of En Route Real-Time Travel Time Information*. Washington, DC: Federal Highway Administration. Available at http://tmcpfs.ops.fhwa.dot.gov/cfprojects/uploaded_files/real_time_tt_rpt.pdf

Liang, Y., Lee, J. D., & Yekhshatyan, L. (2012, December). How Dangerous Is Looking Away From the Road? Algorithms Predict Crash Risk From Glance Patterns in Naturalistic Driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society 54* (6), pg. 1104-1116.

Perez, M., Owens, J., Viita, D., Angell, L., Ranney, T., Baldwin, S., & Parmer, E.. (2012). *Summary of Radio Tuning Effects on Visual and Driving Performance Measures – Simulator and Test Track Studies*. Washington, D.C.: National Highway Traffic Safety Administration. (Later published as Perez, M., Owens, J., Viita, D., Angell, A., Ranney, T. A., Baldwin, G. H. S., Parmer, E., Martin, J., Garrott, W. R., & Mazzae, E. N. (2013, May). *Radio tuning effects on visual and driving performance – Simulator and test track studies.* (Report No. DOT HS 811 781). Washington, DC: National Highway Traffic Safety Administration. Available at www.nhtsa.gov/DOT/NHTSA/NVS/Vehicle%20Research%20&%20Test%20Center%20(VRTC)/ca/capubs/811781.pdf)

Ranney, T., Baldwin, S., Parmer, E., Martin, J., & Mazzae, E. (2011a). *Developing a Test to Measure Distraction Potential of In-Vehicle System Tasks in Production Vehicles*. (Report No. DOT HS 811 463). Washington, DC: National Highway Traffic Safety Administration. Available at www.distraction.gov/download/research-pdf/811463v1.pdf

Ranney, T., Baldwin, S., Parmer, E., Martin, J., & Mazzae, E. (2011b). *Distraction Effects of Manual Number and Text Entry While Driving*. (Report No. DOT HS 811 510). Washington, DC: National Highway Traffic Safety Administration. Available at www.distraction.gov/download/research-pdf/811510v508.pdf

Ranney, T., Baldwin, S., Parmer, E., Martin, J., & Mazzae, E. (2012, February). *Distraction Effects of In-Vehicle Tasks Requiring Number and Text Entry Using Auto Alliance's Principal 2.1B Verification Procedure*. (Report No. DOT HS 811 571). Washington, DC: Available at www.distraction.gov/download/research-pdf/Distraction_Effects_of_In-Vehicle_Tasks_508.pdf

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422.

Reed, N., & Robbins, R. (2008). *The effect of text messaging on driver behavior. A simulator study* (Report No. PPR 367). Berkshire, UK: Transport Research Laboratory.

Regan, M. A., Lee, J. D., & Young, K. L. (2008). *Driver Distraction: Theory, Effects and Mitigation*. Boca Raton, FL: CRC Press.

Reimer, B., Mehler, B., & Coughlin, J. F. (2012). *An Evaluation of Typeface Design in a Text-Rich Automotive User Interface*. (White Paper 2012-12). Cambridge, MA: MIT AgeLab. Available at http://agelab.mit.edu/files/AgeLab_typeface_white_paper_2012.pdf

Rockwell, T. (1988). Spare visual capacity in driving-revisited: New empirical results for an old idea. In A. G. Gale, M. H. Freeman, C. M. Haslegrave, P. Smith, and S. P. Taylor, (eds.), *Vision in Vehicles II. Proceedings of Second International Conference on Vision in Vehicles*, 317-324, Nottingham, UK, September 14-17, 1987. Burlington, MA: Elsevier.

Society of Automotive Engineers. (1998). *SAE Recommended Practice for Calculating the Time to Complete In-Vehicle Navigation and Route Guidance Tasks (SAE J2365). Committee Draft*. Warrendale, PA: Author.

Shannon, C. E. (1951). Prediction and entropy of printed English. *Bell System Technical Journal, 30*(1), 50-64.

Stevens, A., Quimby, A., Board, A., Kersloot, T., & Burns, P. (2002). *Design guidelines for safety of in-vehicle information systems*. Berkshire, UK: Transportation Research Laboratory.

Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices, Notice of Proposed Federal Guidelines (pp. 11199 -11250): Federal Register 77:37 (proposed February 24, 2012).

Wierwille, W. W. (1993). Demands on driver resources associated with introducing advanced technology into the vehicle. *Transportation Research Part C, 1*(2), 133-142.

Wierwille, W. W., Antin, J., Dingus, T., & Hulse, M. (1988). Visual Attentional Demand of an In-Car Navigation Display System. In A. G. Gale, M. H. Freeman, C. M. Haslegrave, P. Smith, and S. P. Taylor,(eds.), Vision in Vehicles II, *Proceedings of Second International Conference on Vision in Vehicles.* (pp. 307-316). Nottingham, UK, September 14-17, 1987. Burlington, MA: Elsevier.

Yavuz, D. (1974). Zipf's law and entropy (correspondence). In Helmut Bölcskei (ed.), *IEEE Transactions on Information Theory, 20*(5), 650-650.

DOT HS 811 820
August 2013

U.S. Department
of Transportation

**National Highway
Traffic Safety
Administration**

NHTSA
www.nhtsa.gov

9749-081313-v2