Author(s): Grahn, Hilkka; Kujala, Tuomo

Title: Impacts of Touch Screen Size, User Interface Design, and Subtask Boundaries on In-Car Task's Visual Demand and Driver Distraction

Year: 2020

Version: Accepted version (Final draft)

Please cite the original version:

Grahn, H., & Kujala, T. (2020). Impacts of Touch Screen Size, User Interface Design, and Subtask Boundaries on In-Car Task's Visual Demand and Driver Distraction. International Journal of Human-Computer Studies, 142, Article 102467. https://doi.org/10.1016/j.ijhcs.2020.102467

Journal Pre-proof

Impacts of Touch Screen Size, User Interface Design, and Subtask Boundaries on In-Car Task's Visual Demand and Driver Distraction

Hilkka Grahn , Tuomo Kujala

PII: S1071-5819(20)30069-0
DOI: https://doi.org/10.1016/j.ijhcs.2020.102467
Reference: YIJHC 102467

To appear in: *International Journal of Human-Computer Studies*

Received date: 27 May 2019
Revised date: 6 March 2020
Accepted date: 13 May 2020

Please cite this article as: Hilkka Grahn , Tuomo Kujala , Impacts of Touch Screen Size, User Interface Design, and Subtask Boundaries on In-Car Task's Visual Demand and Driver Distraction, *International Journal of Human-Computer Studies* (2020), doi: https://doi.org/10.1016/j.ijhcs.2020.102467

Impacts of Touch Screen Size, User Interface Design, and Subtask Boundaries on In-Car Task's Visual Demand and Driver Distraction

Hilkka Grahn*[1], Tuomo Kujala[1]

* Corresponding author

[1]University of Jyväskylä

P.O. Box 35

FI-40014 University of Jyväskylä

Finland

E-mail: hilkka.grahn@jyu.fi, tuomo.kujala@jyu.fi

ABSTRACT

Visual distraction by secondary in-car tasks is a major contributing factor in traffic incidents. In-car user interface design may mitigate these negative effects but to accomplish this, design factors' visual distraction potential should be better understood. The effects of touch screen size, user interface design, and subtask boundaries on in-car task's visual demand and visual distraction potential were studied in two driving simulator experiments with 48 participants. Multilevel modeling was utilized to control the visual demands of driving and individual differences on in-car glance durations. The 2.5" larger touch screen slightly decreased the in-car glance durations and had a diminishing impact on both visual demand and visual distraction potential of the secondary task. Larger relative impact was discovered concerning user interface design: an automotive-targeted application decreased the visual demand and visual distraction potential of the in-car tasks compared to the use of regular smartphone applications. Also, impact of subtask boundaries was discovered: increase in the preferred number of visual or visual-manual interaction steps during a single in-car glance (e.g., pressing one button vs. typing one word) increased the duration of the in-car glance and its visual distraction potential. The findings also emphasize that even if increasing visual demand of a task – as measured by in-car glance duration or number of glances – may increase its visual distraction potential, these two are not necessarily equal.

Keywords: Visual distraction; visual demand; visual occlusion; in-vehicle user interface; subtask boundary; multilevel model

1 INTRODUCTION

Over the past decade, the effects of smartphone usage on traffic safety have been in the focus of research all over the world. Several studies have explored the detrimental impacts of using smartphone while driving with different methods. For instance, there have been many naturalistic (e.g., Guo et al., 2010; Tivesten and Dozza, 2015) and simulator studies (e.g., Choudhary and Velaga, 2017; He et al., 2015a; Rumschlag et al., 2015) as well as surveys (e.g., Bayer and Campbell, 2012; Gauld et al., 2017) and meta-analyses (e.g., Caird et al., 2014; Oviedo-

Trespalacios et al., 2016) investigating drivers' smartphone use and its effects on driver performance and safety. As a general finding, these studies have strengthened the association between smartphone usage and drivers' visual distraction.

One countermeasure to mitigate the distraction smartphone usage causes to drivers is through legislation. For instance, several states in the US have forbidden the usage of cellular phones while driving and almost each state have posited texting bans. However, the use of smartphones while driving goes beyond just texting, since drivers tend to use all kinds of phone applications – from Facebook to YouTube (Ahlström et al., 2019; Kujala and Mäkelä, 2018).

Unfortunately, the user interfaces (UI) of regular smartphone applications are rarely designed to be visually and cognitively low demanding for a car driver. The lack of driver-friendly user interfaces for these applications raises a need for in-car systems that are optimized for the automotive context and which can provide easy access to information and entertainment drivers need on the road. If designed well and accepted by drivers, these interfaces could diminish drivers' visual distraction as well as the use of smartphone applications while driving since the legislation has not fulfilled this urge (e.g., Gauld et al., 2017). However, little is still known about the exact user interface design factors which can effectively reduce drivers' visual distraction by secondary activities.

That said, one design factor that could mitigate drivers' visual distraction is to utilize speech-to-text function in order to decrease the visual-manual demands of an in-vehicle system. Speech-to-text technology recognizes driver's speech and converts it into commands that the system can understand. Various studies have assessed the efficacy of speech-to-text function (or voice recognition) to mitigate driver distraction compared to manual text entry (e.g., Beckers et al., 2017; He et al., 2015b, 2014; Tsimhoni et al., 2004). Typically, manual text entries are nowadays conducted with touch screen keyboards – which are visually highly demanding and causing distraction for drivers (e.g., Crandall and Chaparro, 2012; Kujala et al., 2013; Kujala and Grahn, 2017; McKeever et al., 2013; Reimer et al., 2014b). Another design factor that could mitigate drivers' visual distraction is to utilize read-aloud function which is a technology that reads selected text aloud. However, there is little published glance duration data on read-aloud function in the driving context. According to Owens et al. (2011) read-aloud function does not cause longer glances away from the road compared to baseline driving. Conversely, other studies that are not based on glance duration metrics have found that read-aloud functions may not be distraction-free either (Jamson et al., 2004; Lee et al., 2001). Actually, Reimer and Mehler (2013) and Reimer et al. (2014a) have pointed out that both speech-to-text and read-aloud functions are rarely completely free of visual or manual interaction and can therefore be visually distracting. There are also some task types, such as checking all nearby gas stations from a navigation system, that may be inefficient to be conducted using verbal communications only as the read-aloud function would orally list all the options.

Previous research has also established, in general, that screen size affects efficiency when conducting different tasks (Hancock et al., 2015; Raptis et al., 2013). Gaffar and Kouchak (2017) studied in automotive context drivers' reaction times when selecting target icon on either 7" or

10" screen. They concluded that there was no difference in reaction times between those two relatively large screens. No glance durations were measured in their study. Hence, to our best knowledge, there are no existing well-controlled studies in automotive context about the effects of screen size on glance durations, comparing for instance a smartphone screen versus a tablet screen. Similarly, previous studies have not extensively dealt with screen orientation's effects on visual distraction in detail.

Yet another design factor that could diminish drivers' visual distraction while conducting secondary in-car tasks, are well designed task structures (i.e., "how a task breaks down into smaller subtasks" [Salvucci and Kujala, 2016]) that are based on scientific knowledge of human multitasking behavior. It has previously been observed that people have a tendency to switch tasks at subtask boundaries (e.g., Janssen et al., 2012; Lee et al., 2015; Lee and Lee, 2019; Salvucci and Kujala, 2016), for instance dialing a phone number in chunks or typing one word at a time (Janssen et al., 2012). Empirical evidence also suggests that when the duration of a secondary visual search task increases, the glance durations tend to increase as well (Kujala and Salvucci, 2015; Lee et al., 2012). Particularly in a time-critical situation like driving, these findings are crucial to take into account when designing user-interfaces for the automotive context.

In order to clarify some of the key design factors which may have an impact on drivers' visual distraction, we studied features of a novel automotive-targeted infotainment application called Carrio – which is designed exclusively for in-car use. Since drivers use smartphones while driving for various means, we compared Carrio's visual distraction potential to regular (Android) smartphone applications and studied different in-car tasks' visual demands in two experiments with 48 participants in a driving simulator. Our general intention was to examine if – and how much – an automotive-targeted application is able to reduce real-world in-car tasks' visual distraction potential compared to regular smartphone applications. The research questions were:

RQ1) Are there significant differences in the visual distraction potential between automotive-targeted application (Carrio) and regular smartphone applications (Android)?
RQ2) Are there differences in the visual demand of the tasks conducted with automotive-targeted application (Carrio) and regular smartphone applications (Android)?
RQ3) If there are differences, what are the design factors that are responsible for these?

We have divided the remaining paper into seven sections The second section after the introduction describes the general method used in both experiments. The third section deals with Experiment 1 studying visual distraction potential of automotive-targeted application compared to regular smartphone applications. Since Experiment 1 did not fully answer the posited research questions, we conducted second experiment, which is dealt in the fourth section of the paper. Experiment 2 examines visual distraction potential of the same applications as in Experiment 1 while the effects of screen size and screen orientation are controlled for. The fifth section presents two multilevel models constructed using data from both experiments. These multilevel models enable studying the effects of screen size, screen orientation, application, and task type on in-car glance durations together when controlling for visual demands of driving and individual differences between the participants. The sixth section deals with general discussion and answers

to the posited research questions. Lastly, the seventh section summarizes the conclusions of this paper.

Results of Experiment 1 indicated that an automotive-targeted application (Carrio) seemed to diminish visual distraction compared to regular smartphone applications. Since larger screen size and landscape orientation could have favored this automotive-targeted application, we conducted Experiment 2 to control for these possible effects. Results of Experiment 2 indicated that only one (email replying task) out of three tasks conducted with the automotive-targeted application decreased visual distraction significantly compared to regular smartphone applications. In order to further analyze the combined data of the two experiments while controlling for visual demands of driving and individual differences, we used multilevel modeling to study how screen size, screen orientation, application, and task type affect visual demand of the tasks.

Utilizing data from both experiments, we constructed two multilevel models. Based on the multilevel models, the 2.5" larger screen slightly decreased the in-car glance durations and thus, diminished the visual demand of the in-car tasks. However, the type of application had larger relative impact than the screen size – the automotive-targeted application (Carrio) seemed to decrease the visual demand and visual distraction potential of the in-car tasks compared to regular smartphone applications (Android). A possible impact of subtask boundaries were also recognized: driver's ability to break in-car tasks into smaller subtasks seem to decline individual in-car glances' durations, facilitating better adjustment of glancing behavior in relation to the demands of the driving situation. Additionally, the findings of the paper stress that visual demand of a task – measured as in-car glance duration or number of glances – and visual distraction potential of the task are not inevitably equal.

2 GENERAL METHOD

For measuring the visual distraction potential of different in-car tasks, we used a method introduced by Kujala and Mäkelä (2015), which has been applied in studies by Grahn and Kujala (2018), Kujala et al. (2016a) as well as in Kujala and Grahn (2017). The method contains two parts: visual distraction potential testing and driver sample validation. This novel method was used in order to categorize the in-vehicle glances as appropriate or inappropriate based on the situational visual demand of the driving task, to control these demands of the driving task in statistical modeling, and to control the driver sample.

The method of Kujala and Mäkelä (2015) utilizes visual occlusion technique, which was initially introduced by Senders et al. (1967). Traditionally visual occlusion refers to a condition where the driver's vision is occasionally occluded (i.e., driving blind) and the *duration* of the self-selected occlusion is measured. In the method we used, *distance* that is driven during the occluded period is measured, not *time* as Senders et al. (1967) did. This enables free speed control for the driver. Here, the blindly driven distance is called occlusion distance (OD). Higher OD can be interpreted to refer to lower visual demands of driving.

In the visual distraction potential testing, the operationalization of visual distraction is based on the data collected by Kujala et al. (2016b), where 97 drivers' preferred occlusion distances on

simulated highway and suburban roads were measured while the drivers were only focusing on safe driving. Afterwards the measured occlusion distances were mapped on the same test routes. This means that each 1x1 meter route point in the map holds information on occlusion distances that were driven in that particular route point in the experiment. When the same test routes are used in the visual distraction potential testing, it enables the categorization of in-car glances (i.e., glances that are directed to the in-car device) to appropriate or inappropriate. The categorization is based on the distance driven during the in-car glance from a particular route point where the glance begins. An inappropriate or *red in-car glance* refers to an in-car glance length that exceeds the 85th percentile of the 97-driver sample's occlusion distance on the route point. Red glances can thus be considered as inappropriately long in-car glances in relation to the visual demand of the given driving situation and which can be considered as visual distraction.

This operationalization of visual distraction takes into account the dynamic visual demands of the driving task. Compared to operationalizations of visual distraction with static glance thresholds (e.g., 2.0 seconds by NHTSA, 2013), it enables the driver more tactical freedom to adjust these demands by lowering speed and/or selecting low-demand conditions for interacting with in-car devices.

Visual demand of the secondary in-car task is often operationalized as mean or total in-car glance durations (e.g., NHTSA, 2013) or as number of in-car glances but the effects of the variable visual demands of the driving task on these are not considered. For instance, Wierwille (1993) has shown that the demands of the driving task affect significantly off-road glance durations. Measurement of the visual demand of the driving situation as the median-OD of the 97-driver sample enables control of this variable factor in the statistical modeling of the visual demand of the in-car tasks as in-car glance durations.

In addition, in the method, occlusion distances are utilized to validate the driver sample to match the occlusion distance distribution with the 97-driver sample (Kujala et al., 2016b) to ensure that the sample contains all kinds of drivers – from those who are able to drive short distances occluded to those who are able to drive long distances occluded. That is, participants who need much visual information on the road and participants who need less visual information on the road to be able to drive safely and accurately. This is important, since according to previous studies, drivers tend to have individual preferences for off-road glance durations (Broström et al., 2016; Kujala et al., 2014), which may have an effect on the distraction test results (Broström et al., 2013; Lee and Lee, 2017).

3 EXPERIMENT 1

In Experiment 1 we wanted to study if there are significant differences between the distraction potential of an automotive-targeted application (Carrio) and regular smartphone applications (Android). To study this, 24 participants conducted common in-car tasks during simulated driving with Carrio and Android applications.

3.1 Method

3.1.1 Experimental design

The experimental design for the analyses of the effects of the in-car task types between automotive-targeted application (Carrio) and regular smartphone applications (Android) was within-subjects 2 x 3. The independent variables (IV) were application (Carrio and Android) and task type (email reading, view-switching, song searching). The dependent variables were number of in-car glances and percentages of red in-car glances.

3.1.2 Participants

In all, 24 participants were recruited by convenience sampling using different mailing lists. The NHTSA (2013) recommendations on the driver sample selection were followed as accurately as possible. Seventeen participants were male and seven were female. The imbalance between the genders was a result from simulator sickness: five females with symptoms were replaced with males.

Eight participants were 18 to 24 years old, nine 25 to 39 years old, four 40 to 54 years old and three were older than 55 years. The age of participants varied from 20 to 79 years, mean age being 34.8 years (*SD* = 16.0). Each participant had a valid driver's license and drove at least 5 000 kilometers per year. The driven kilometers per year varied from 5 000 to 30 000 kilometers with a mean of 12 938 kilometers (*SD* = 7 046) per year. The range of driving experience was from two to 55 years and the mean was 16.0 years (*SD* = 15.0). All participants were generally healthy and had normal or corrected-to-normal vision. The experiments were instructed in Finnish and all participants understood and spoke Finnish. Time to complete the experiment ranged from 1 hour and 7 minutes to 1 hour and 40 minutes. After the experiment, each participant was rewarded with a gift certificate (15 EUR).

3.1.3 Apparatus

The experiments took place at the University of Jyväskylä's driving simulator laboratory. The medium-fidelity driving simulator has a CKAS Mechatronics 2-DOF motion platform, automatic transmission, longitudinally adjustable seat as well as Logitech G27 force-feedback steering wheel and pedals. The simulator had three 40" LED screens (95.6 cm x 57.4) with a resolution of 1440 x 900 pixels per screen. The middle screen displayed a head-up display (HUD) tachometer, a HUD speedometer, and a rear-view mirror, and the side screens had side mirrors (see Figure 1).

Figure 1: The experimental setup. The smartphone and the tablet are located next to the steering wheel. The participant is wearing a head-mounted eye-tracker.

Driving simulator software was provided by Eepsoft (http://www.eepsoft.fi/) and it saved the driving log data at 10 Hz. The steering wheel was outfitted with two levers that exposed the driving scene for 500 milliseconds per press during the occlusion trial following the original occlusion method by Senders et al. (1967). If the levers were constantly pressed, the driving scene was constantly visible. The routes used simulated real suburban roads that are located in southern Finland. The routes were copied from the study of Kujala et al. (2016b).

Ergoneers' Dikablis Essential 50 Hz head-mounted eye-tracking system was used to record eye movements. To synchronize the driving simulator data (x, y, and speed) and the eye-tracking data, a LAN bridge and a custom logging software were used.

The automotive-targeted application (Carrio, https://carrioapp.com/) was running on a 7" Lenovo TB3-730X tablet (Android 6.0). A Samsung Galaxy A3 smartphone (4.5", Android 6.0.1) was utilized to run different applications that were compared to the Carrio application (see Figures 2 and 3). Both devices were on a holder placed on the right side of the steering wheel (see Figure 1). Carrio was used in landscape mode for which the application is optimized for, whereas the smartphone was in portrait mode, which can be argued to be the most typical mode of use for smartphones in this context (i.e., for single-handed use) and the Android operating system's default mode. Rstudio (version 1.0.136) and IBM SPSS Statistics 24 were utilized to conduct the statistical analyses.

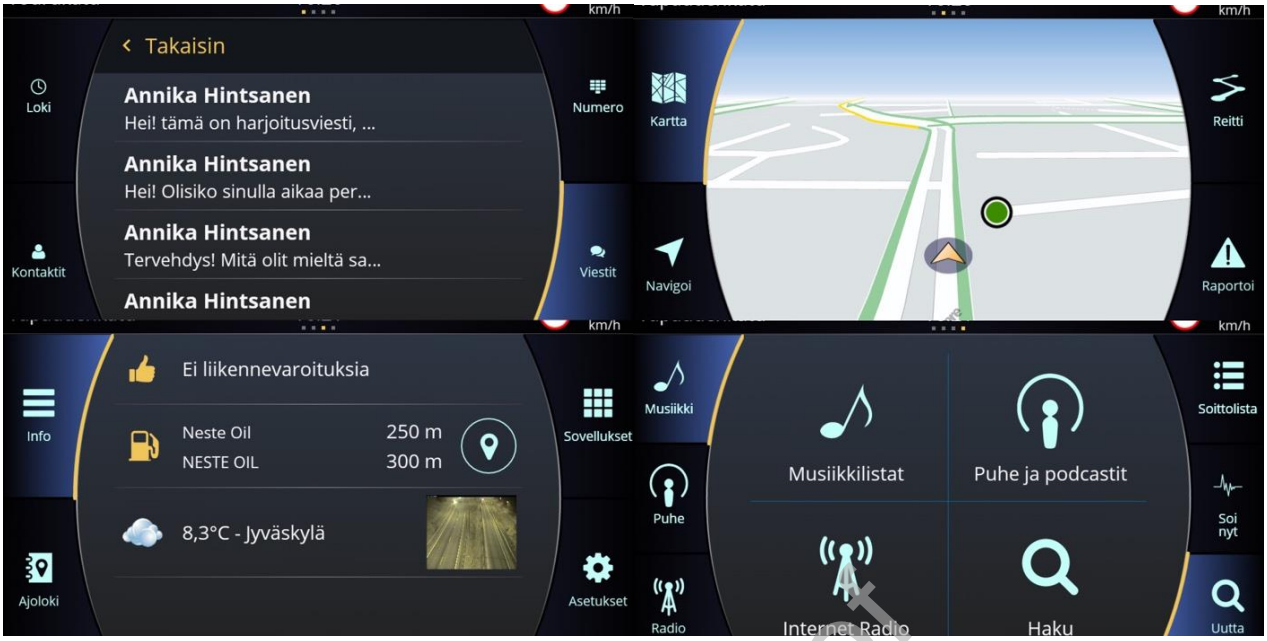

Figure 2: Views of the automotive-targeted application (Carrio). Left upper corner: email reading view, right upper corner: navigation view, left lower corner: weather condition view, and right lower: corner song searching view. The navigation view and the weather condition view were only used in the view switching task.

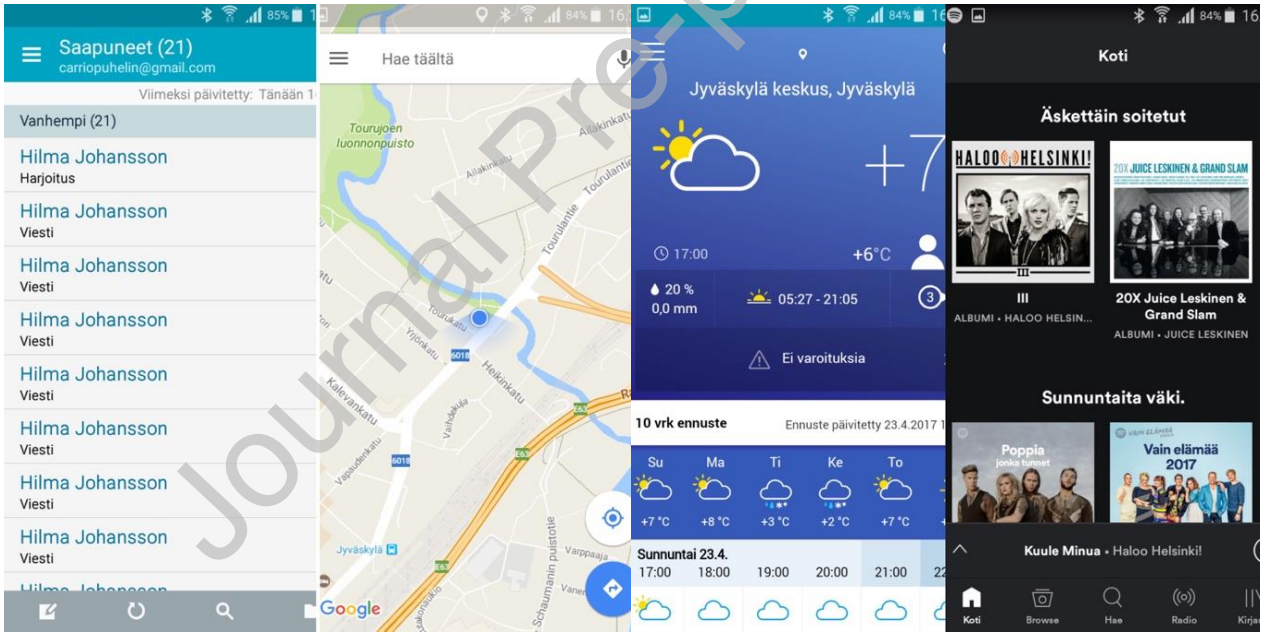

Figure 3: Views of regular smartphone applications (Android). From left: email reading view, navigation view, weather condition view, and song searching view. The navigation view and the weather condition view were only used in the view switching task.

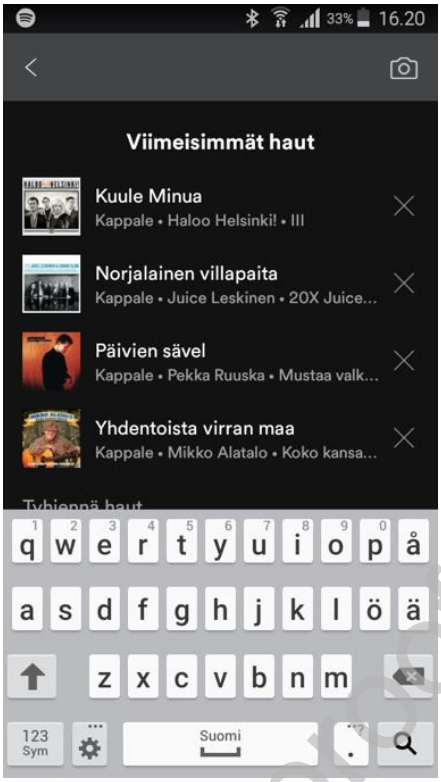

Figure 4: The qwerty touch screen keyboard of the Android smartphone.

3.1.4 Procedure

Demographic data was collected beforehand via email. Upon arrival, participants read and signed the informed consent form and were informed about the purpose and the setup of the study. First participants practiced driving with the driving simulator in an artificial city environment with other traffic. This was done in order to gain experience on driving the simulator, especially on left and right turns. The average practice time was 6 minutes. After they felt comfortable with driving, they started to practice for the occlusion trial, that is, how to drive when vision is occasionally occluded. The environment for the occlusion trial was the same artificial city with other road users as in the previous practice but the starting point was different. The average practice time was four and half minutes.

First task after the practices was the occlusion trial for the validation of the test sample. During the occlusion trial the driving simulator's screens were blank by default and the driving scenery could be revealed for 500 milliseconds (as in Senders et al., 1967) by pressing the levers that were attached to the steering wheel. The two-lane highway route without traffic was the same that was used to gather the baseline data of Kujala et al. (2016b). Before the trial, each participant received instructions to obey the traffic regulations, to drive safely and accurately but still to try to drive without visual information (i.e., vision occluded) as long as possible. Those six participants who could drive the longest median distances without visual information and still accurately, were promised a movie ticket as an extra reward. The reward was promised in order to get participants to concentrate on the driving task but still trying to maximize the occlusion distance to their preference. The speed limits varied from 60 to 80 to 120 kilometers per hour depending on the highway section. Every change in the speed limit was told to each participant at the same point of

the route. However, sections that were driven 60 km/h were junctions from a highway to another highway and were not included in the final data. After the trial, NASA-TLX questionnaire (Hart and Staveland, 1988) was filled out in order to measure task workload.

The distraction potential testing followed the occlusion trial. The routes used were the same suburban roads as in Kujala et al. (2016b). First, the head-mounted eye-tracker was put on, adjusted and calibrated. The instructed speed limit during the trials was 50 kilometers per hour but the speed could be adjusted freely if needed. Before each task the experimenter gave instructions and showed how to perform similar tasks as in the actual distraction testing. The participants repeated the example tasks to get familiar with the features of the Carrio application and the different applications of the Android smartphone.

In the beginning of the distraction potential testing, the participants were told to prioritize driving, to follow traffic regulations, and to drive safely. Each participant conducted three different types of tasks per application (Carrio and Android). After every task type, there was a short break with instructions to the following task. The task types were selected to represent plausible activities drivers may be conducting with their smartphones while driving, related to information search, entertainment, and switching between applications. The tasks are listed in Table 1.

Experiment 1	Email reading	View-switching	Song searching
	Read 20 emails (104–179 characters in one email) and search answers to questions asked by the experimenter. In total four questions.	Switch between different views (email, map, weather, Spotify). 15 switches in total.	1) Search and start to play a song announced by the experimenter (repeated four times). 2) Look for the album where the song in question is included (repeated two times) OR look for the five most popular songs of the artist who performed the song in question (repeated two times).

Table 1: Tasks in Experiment 1.

The task procedures differed between the applications. Task procedures are explained in Table 2.

Experiment 1	Email reading (20 tasks)	View-switching (15 tasks)	Song searching (4 tasks)
Carrio	Application read the selected emails out loud (read-aloud function) by tapping the message header. One button press per task.	Conducted by swiping the screen either to left or right to find the right view. One to three swipes per task.	Conducted using speech-to-text function for searching the target songs. Found songs were selected by tapping the right one from a list of suggestions. A song was associated with a menu in which the information about the album or the artist could be opened by tapping the desired menu item. Nine button presses per task.
Android	Participants read the emails by themselves by tapping message headers one by one and then returning back to the list of received emails. Two button presses per task.	Conducted by pressing a button on the left lower corner of the phone that presented all the active applications and the right one was chosen by tapping it. Two button presses per task.	Conducted using a keyboard for searching the target songs and by selecting the correct items on the menus associated with the playing song. In addition to typing the names of the songs, seven button presses per task were required.

Table 2: Task procedures in Experiment 1.

Three different routes were used (see Figure 5) and the order of the routes and the tasks were counterbalanced. However, similar tasks between the applications were always done on the same routes per participant. The same routes for equivalent tasks were used in order to control the visual demands of the routes. There was no other traffic on the roads during the trials. After each task, NASA-TLX questionnaire (Hart and Staveland, 1988) was filled out, in total six times during the distraction testing.

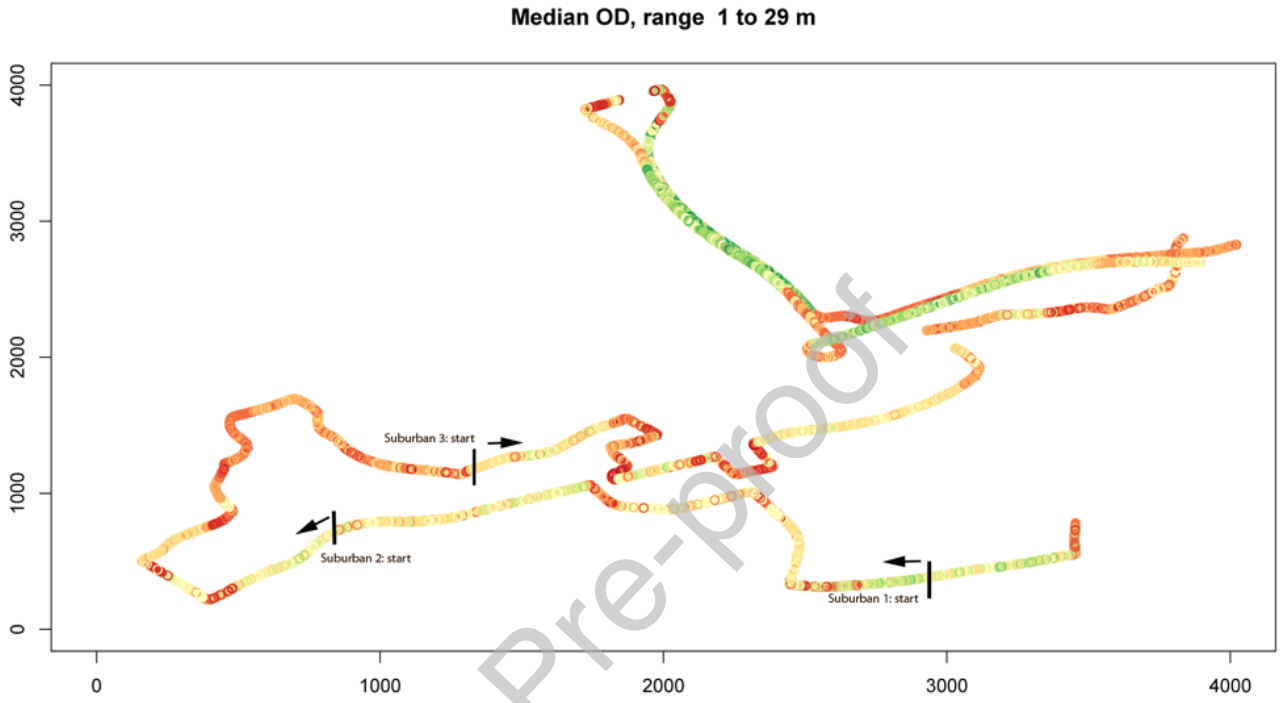

Figure 5: The pre-defined routes for the experiments. Color indicates the visual demand of the route point as occlusion distance: the demand increases from green to yellow to orange to red.

3.1.5 Data preparation

For measuring the occlusion distances, driving simulator's log data was used to automatically calculate the driven distance during the occlusion. A script calculated and logged the driven distance in between the start of each occlusion event and the following lever press based on the odometer reading. Scoring of the in-car glance lengths was conducted in real time with a script that read the x and y coordinates of the pupil as well as the timestamp provided by the eye-tracker. The pupil coordinates were synchronized with the location and timestamp data afforded by the driving simulator. After the experiment, Noldus Observer XT software was used to manually inspect each in-car glance length from a synchronized video (25 frames per second) provided by the eye-tracking software. All inaccuracies were manually corrected frame-by-frame. SAE-J2396 (Society of Automotive Engineers, 2000) definition was followed when scoring the in-car glance lengths. However, to enable more direct comparability with the occlusion distance, the gaze transition time back to the driving scene from the in-car device was added to the glance length. Based on the data by Kujala et al. (2016a), all in-car glances exceeding the 85th percentile of the original 97-driver sample's occlusion distance on the route point were categorized as red in-car glances.

Unfortunately, due to technical issues during the trials, and thus missing data points, one participant had to be removed from the objective data (*N* = 23). However, raw NASA-TLX questionnaire (Hart, 2006; Hart and Staveland, 1988) results, that are reported in Table 11, include all participants (*N* = 24).

3.2 Results and discussion

3.2.1 Occlusion distances

To validate the driver sample, the distribution of the occlusion distances was compared to the original occlusion distance distribution of the 97-driver sample (Kujala et al., 2016b) where the occlusion distances varied from 3.21 meters to 41.88 meters (*Mdn* = 13.67). In this experiment, the occlusion distances varied from 6.35 meters to 35.82 meters with a median of 17.37 meters. According to Levene's test, the variance of the occlusion distance distribution does not differ significantly from the original OD distribution of Kujala et al. (2016b): (F (1,116) = .645, p = .424).

3.2.2 Number of in-car glances by user interface and task type

The number of in-car glances for each task type was sufficient for meaningful and reliable analyses, see Table 3. According to paired-samples t-test, differences between email reading task (t(22) = -7.188, p < .001, d = 1.895) and view-switching task (t(22) = -10.642, p < .001, d = 2.340) were significant, Carrio having lower mean number of in-car glances. No significant difference was found in song searching task (p = .468).

Experiment 1	Email reading (20 tasks)	View-switching (15 tasks)	Song searching (4 tasks)
Carrio	*M* = 41.52 (12.43)	*M* = 20.12 (9.58)	*M* = 65.14 (19.11)
Android	*M* = 86.83 (31.44)	*M* = 39.57 (6.81)	*M* = 63.78 (23.68)

Table 3: Mean number of in-car glances per user interface task type (standard deviation in parentheses).

3.2.3 Red in-car glance percentages by user interface and task type

Because the distributions of the red in-car glance percentages were non-Gaussian, median was used as a measure of central tendency instead of mean in statistical testing. According to Wilcoxon Signed Rank test, all the differences in the percentages of red in-car glances between Carrio and Android applications per task type were significant with medium effect sizes in favor of Carrio (email reading: Z = 2.584, p = .010, d = 0.666; view-switching: Z = 2.458, p = .014, d = 0.688; song searching: Z = 2.795, p = .005, d = 0.677, see Table 4).

Experiment 1	Email reading	View-switching	Song searching
Carrio	*Mdn* = 10.00 (16.00)	*Mdn* = .00 (3.25)	*Mdn* = 6.00 (13.50)
Android	*Mdn* = 19.00 (19.50)	*Mdn* = 6.00 (8.25)	*Mdn* = 16.00 (14.25)

Table 4: Median red in-car glance percentages per user interface and task type (interquartile range in parentheses).

3.2.4 Discussion

In Experiment 1, based on the significant differences on red in-car glances (i.e., visual distraction), it can be argued that the tested features of Carrio application seem to have significantly lower visual distraction potential compared to similar tasks conducted with regular smartphone applications (Android) while driving.

The tested applications had some differences in their functionalities which are reported in Table 2. Android's email reading task had the highest percentages of red in-car glances (19.00 %). This task did not demand many button presses (2 per email) or typing but it demanded reading from the screen in order to complete the task. When conducted with Carrio application (red in-car glance: 10.00 %), the task required one button press per email and no reading since the application read the email aloud. Second highest percentage of red in-car glances (16.00 %) was discovered in the Android smartphone's song searching task. This task required several button presses (7 per song) and some typing with quite small buttons (see Figure 4) before the target song was found. When conducted with Carrio application (6.00 %), this task required nine button presses and no typing as the application utilized speech-to-text function.

The view-switching task with both applications had the lowest percentages of red in-car glances (.00 % vs. 6.00 %). Conducted with Android smartphone, this task required two button presses per switch. Carrio's view-switching task required only swiping between four different views (one to three swipes per task) and participants were able to learn easily the order of the views. This may have enabled participants to keep their eyes on the road during the task. On the other hand, Android's view-switching task required more looking at the device since the participants had to tap to the right spot on the screen in order to complete the task. Overall, Carrio application had lower distraction potential compared to similar tasks conducted with regular smartphone applications.

During the experiment, both devices were placed in a holder (see Figure 1) that was moving with the simulator. The motion of the driving simulator may have increased the difficulty of conducting the tasks but it may have affected the Carrio tasks less because of application's larger buttons and the screen size of the tablet.

Overall, these findings indicate the potential of well-designed in-car applications to decrease visual distraction compared to the use of regular smartphone applications while driving. The automotive-targeted application Carrio is designed to be used easily while driving with its large buttons, multimodal interactions, and simplistic design. We assume that the advantage of Carrio in the tested tasks could have been mainly due to the speech-to-text function in the song searching tasks and read-aloud functions in the email tasks. In addition, Carrio was used in a tablet with larger screen and in landscape mode whereas the smartphone had 2.5" smaller screen and was used in portrait mode – this arrangement could have also favored Carrio.

Due to these confounding factors, we cannot exactly pinpoint with this data alone the most important design factors that favored Carrio. Therefore, we cannot fully answer to the posited research questions. To control for these confounded factors, we conducted a second experiment to better answer our research questions.

4 EXPERIMENT 2

Our aim in Experiment 2 was to examine if the observed lower distraction potential of the Carrio tasks in Experiment 1 was due to the larger tablet screen or its landscape orientation compared to the smaller-sized smartphone in portrait mode. To be able to compare the effects of these factors between the experiments, participants conducted in this experiment the same email reading and

song searching tasks as in Experiment 1. However, this time both Carrio and regular smartphone applications were running on 4.5" smartphone in landscape mode. The view-switching task was omitted since it was found relatively easy and low-distracting in Experiment 1 with both, Carrio and Android. Instead, participants had an extra task to reply to four emails to get data from a text entry task with manual typing compared to a speech-to-text function. Based on previous research (e.g., Crandall and Chaparro, 2012; McKeever et al., 2013; Reimer et al., 2014b), the manual text entry task was assumed to be the most visually distracting due to task structure (preferred typing of word per glance) and manual keyboard input requiring high accuracy.

4.1 Method

4.1.1 Experimental design

Again, the experimental design for the analyses of the effects of task types between automotive-targeted application (Carrio) and regular smartphone applications (Android) was within-subjects 2 x 3 (task types: email reading, email replying, song searching). The independent and dependent variables were the same as in Experiment 1 (IVs: application and task type, DVs: number of in-car glances and percentages of red in-car glances).

4.1.2 Participants

In total 24 participants took part in Experiment 2. None of these participants took part in Experiment 1. The sample was a convenient sample in that regard the participants were recruited via different mailing lists. Again, the NHTSA (2013) recommendations were followed as closely as possible when selecting the participants.

In the driving sample, 16 participants were male and 8 were female. Again, the imbalance between the genders was due to simulator sickness females were reporting. If the participant felt sick during the experiment, the experiment was cancelled and the gathered data was discarded.

Seven participants were 18 to 24 years old, nine 25 to 39 years old, five 40 to 54 years old and three were older than 55 years. The age of participants varied from 19 to 66, mean being 35.3 years (*SD* = 13.9). Each participant had a valid driver's license and drove at least 5 000 kilometers per year. The driven kilometers per year varied from 5 000 to 55 000 kilometers, mean being 14 625 kilometers (*SD* = 11 854) per year. The range of driving experience was from two to 48 years, mean being 16.9 years (*SD* = 13.9). All participants were generally healthy and had normal or corrected-to-normal vision. The experiments were instructed in Finnish and all participants were fluent in Finnish. Time to complete the experiment ranged from 1 hour and 8 minutes to 1 hour and 37 minutes. After the experiment, each participant was rewarded with a gift certificate (10 EUR).

4.1.3 Apparatus

The exactly same driving simulator (see Figure 1), other equipment (excluding the tablet), routes (see Figure 5), and statistical tools were used in this experiment. A software update to the commercial Carrio application between the experiments enabled us to study speech-to-text function in the text entry task. However, an additional smartphone had to be used for this task to keep the same older version of Carrio for the email reading and song searching tasks in the same phone as in Experiment 1. Changing the versions of Carrio during the experiment was evaluated to

be too risky since it could have caused technical difficulties during the experiment. In addition, this would have extended the duration of experiments. Since the same version of Samsung Galaxy A3 did not exist in the market anymore, the additional phone was Samsung Galaxy A3 (2017) with Android 7.0 operating system. It has 4,7” screen which is 0.2” larger than in Samsung Galaxy A3 used in Experiment 1. In each task both smartphones were used in landscape mode in order to be able to control the effects of the screen orientation.

4.1.4 Procedure

Each participant went through exact same practices as in Experiment 1. In the artificial city scenario, the average practice time was five minutes, and in the occlusion drive the average practice time was four minutes. After the practices, the occlusion trial was conducted exactly the same as in Experiment 1. The general instructions, routes, experimenter, counterbalancing, and practicing the mock tasks were the same as in Experiment 1.

Also, two of the selected task types were the same as in the previous experiment – email reading and song searching (see Figures 6 and 7). However, we added an email replying task to compare Carrio’s speech-to-text function to Android’s touch screen keyboard (see Figures 8 and 9). These tasks are listed in Table 5. In addition, NASA-TLX questionnaire (Hart and Staveland, 1988) was filled out after each task, in total seven times.

Figure 6: Views of the automotive-targeted application (Carrio). From left: email reading view and song searching view.

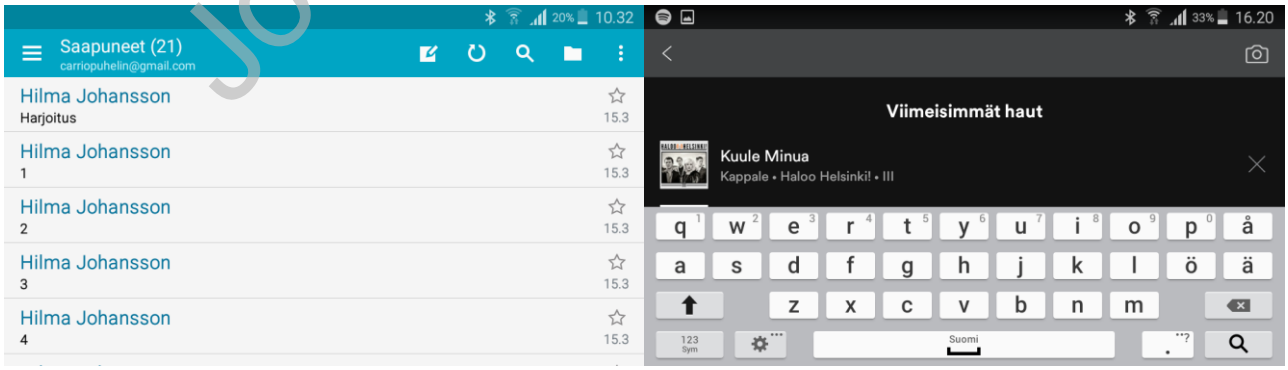

Figure 7: Views of regular smartphone applications (Android). From left: email reading view and song searching view.

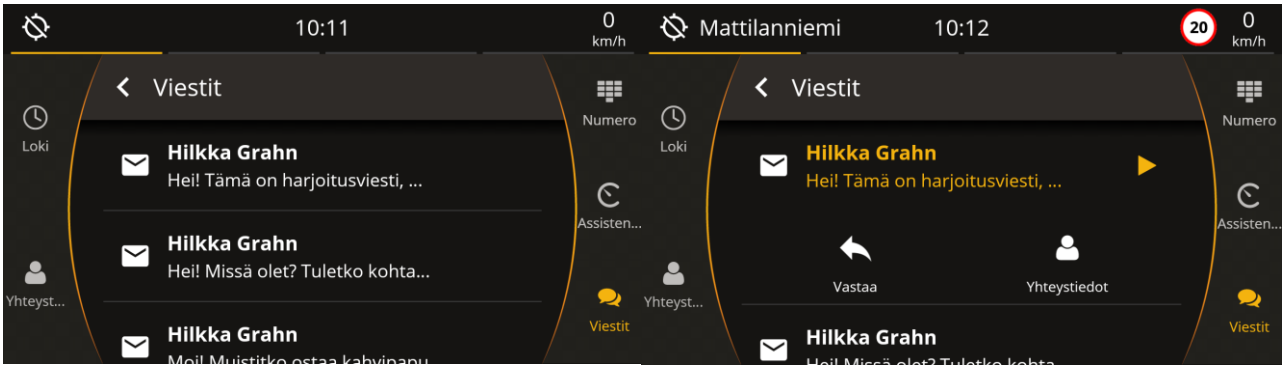

Figure 8: Views of the automotive-targeted application (Carrio). From left: email replying view and view when the email is being listened to.

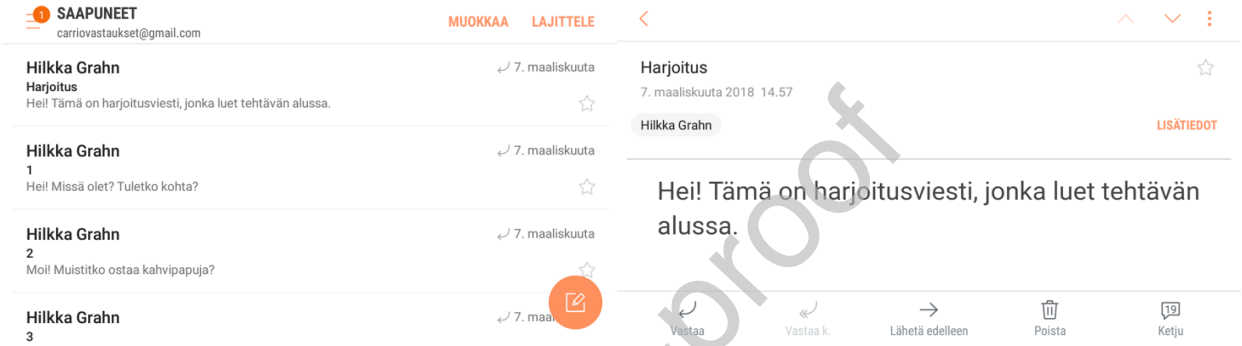

Figure 9: Views of regular smartphone application (Android). From left: email view and view when the email is being replied to.

Experiment 2	Email reading	Email replying	Song searching
	Read 20 emails (104–179 characters in one email) and search answers to questions asked by the experimenter. In total four questions.	Read and reply to emails with an answer told by the experimenter. In total four replies.	1) Search and start to play a song announced by the experimenter (repeated four times). 2) Look for the album where the song in question is included (repeated two times) OR look for the five most popular songs of the artist who performed the song in question (repeated two times).

Table 5: Tasks in Experiment 2.

The task procedures differed a bit depending on the application used. Detailed task procedures can be seen in Table 6.

Experiment 2	Email reading (20 tasks)	Email replying (4 tasks)	Song searching (4 tasks)
Carrio	Application read the selected emails out loud (read-aloud function) by tapping the message header. One button press per task.	Application read the selected emails out loud (read-aloud function) by tapping the message header. Participants replied by tapping a reply button, said the answer out loud and finally tapped send button. Four button presses per task.	Conducted using speech-to-text function for searching the target songs. Found songs were selected by tapping the right one from a list of suggestions. A song was associated with a menu in which the information about the album or the artist could be opened by tapping the desired menu item. Nine button presses per task.
Android	Participants read the emails by themselves tapping message headers one by one and then returning back to the list of received emails. Two button presses per task.	Participants read the emails by themselves, tapped reply button, typed the answer using keyboard (15–23 characters) and finally tapped send button. In addition to typing the replies, four button presses per task were required. Predictive text function was omitted.	Conducted using a keyboard for searching the target songs and by selecting the correct items on the menus associated with the playing song. In addition to typing the names of the songs, seven button presses per task were required.

Table 6: Task procedures in Experiment 2.

4.1.5 Data preparation

The data preparation in this experiment was conducted exactly the same as in Experiment 1.

4.2 Results and discussion

4.2.1 Occlusion distances

Occlusion distances varied from 4.77 meters to 35.99 meters median being 16.53 meters. According to Levene's test, the variance of the occlusion distance distribution does not differ significantly from the original OD distribution of Kujala et al. (2016b): (F (1,117) = .032, p = .859).

4.2.2 Number of in-car glances by user interface and task type

The number of in-car glances for each task type was sufficient for meaningful and reliable analyses (Table 7). According to paired-samples t-test, differences between email reading task (t(23) = -10.028, p < .001, d = 3.020) and email replying task (t(23) = -3.479, p = .002, d = 0.875) were significant, Carrio having lower mean number of in-car glances. No significant difference was found in song searching task (p = .170).

Experiment 2	Email reading (20 tasks)	Email replying (4 tasks)	Song searching (4 tasks)
Carrio	*M* = 38.92 (12.33)	*M* = 38.50 (8.23)	*M* = 57.21 (12.45)
Android	*M* = 85.63 (18.07)	*M* = 51.42 (19.20)	*M* = 51.83 (15.08)

Table 7: Mean number of in-car glances per application and task type (standard deviation in parentheses).

4.2.3 Red in-car glances by user interface and task type

Because the distributions of the red in-car glance percentages were non-Gaussian, median was used as a measure of central tendency instead of mean in statistical testing. According to Wilxocon Signed Rank test, only statistically significant difference was observed between the applications in the email replying task, favoring Carrio (Z = 3.254, p = .001, d = 0.531, see Table 8). Other differences were not significant (email reading: p = .424; song searching: p = .503).

Experiment 2	Email reading	Email replying	Song searching
Carrio	*Mdn* = 14.14 (16.62)	*Mdn* = 4.83 (12.91)	*Mdn* = 8.30 (10.66)
Android	*Mdn* = 13.93 (12.19)	*Mdn* = 12.40 (17.69)	*Mdn* = 8.14 (14.08)

Table 8: Median red in-car glance percentages per application and task type (interquartile range in parentheses).

4.2.4 Discussion

In Experiment 2, based on the significant differences in red in-car glances (i.e., visual distraction), email replying task with Carrio had significantly lower visual distraction potential compared to a similar task conducted with a regular smartphone application (Android) while driving, as hypothesized. However, no significant differences in visual distraction was found in email reading or song searching tasks.

Since in Experiment 2 the tasks were conducted with the same screen size and in landscape mode with both Carrio and Android, Carrio did not have the possible advantage of bigger screen size or orientation as in Experiment 1. This suggests that screen size or device orientation may have a significant impact on visual distraction potential of in-car tasks together with speech-to-text and read-aloud functions.

However, again, due to possible confounding factors, including individual differences between participants, we have to analyze the relative effects of the different UI design factors together with the data from Experiment 1 using multilevel modeling.

5 MULTILEVEL MODELING AND ANALYSES OF DESIGN FACTORS

5.1 Model 1

In order to analyze in detail the effects of screen size, screen orientation, application, and in-car task on in-car glance durations, two multilevel models (Hox, 1998) were created. For Model 1, the glance data from both experiments was organized in a longitudinal format and only tasks which were conducted in both experiments (email reading and song searching) were included in the data. According to the "30/30 rule", sufficient statistical power is reached in multilevel analysis if there are at least 30 observations on level 1 and which are nested on level 2 within 30 units (Richter, 2006). The data contained 11 459 in-car glances (level 1) that belonged to 47 participants (level 2).

5.1.1 Screen size, screen orientation, and application

The dependent variable in the first model was in-car glance duration. The model construction started with exploring the intraclass correlation (ICC) which was 11.89 % in the intercept only model. This justifies the use of a multilevel model. After that, we added fixed factors one by one, inspected the -2 Log-Likelihood Ratio and tested with chi-squared test (χ^2) if the new model had significantly better fit than the previous one. If it had not, the added fixed factor was removed from the model. This inspection revealed that both driving speed and age groups after NHTSA (2013) were significant in the model, that is, affecting glance durations, but they did not significantly improve the fit of the model. Screen orientation had no significant effect on in-car glance duration. We found no significant interactions of the factors.

In the final model (Table 9), as fixed factors we entered user interface, screen size, and occlusion distance. As random factors, we had intercepts for participants (i.e., drivers). After constructing the model, we visually inspected residual plots and they did not indicate any clear deviations from normality or homoscedasticity.

The equation for the first model is:

$$duration_{ij} = b_0 + b_1 size_{ij} + b_2 app_{ij} + b_3 OD_{ij} + u_{0j} + e_{0ij} \qquad (1)$$

where $duration_{ij}$ is in-car glance duration (DV), b_0 is the intercept (grand mean), $b_1 size_{ij}$ is the screen size, $b_2 app_{ij}$ is the application (Carrio or Android), $b_3 OD_{ij}$ is occlusion duration (m, inverse of visual demand of the driving situation), u_{0j} is the random effect (driver), and e_{0ij} is the residual.

Fixed effects	estimate	standard error	*p*	95 % confidence interval
Intercept	932	32	< .001	867–996
Screen size small (4,5”)	39	18	.031	3–75
Screen size large (7”)	0*	0*		
Regular smartphone application (Android)	279	13	< .001	253–304
Automotive-targeted application (Carrio)	0*	0*		
Occlusion distance (m)**	12	1	< .001	10–14
Random effects	σ^2			
Intercept (participant)	35	3	< .001	
Residual	228	7	< .001	
Intraclass correlation (ICC)				
Participant	.133			
Model fit (-2RLL)	15710.22			

Table 9: Multilevel model for in-car glance duration (ms) – Model 1.

* The factor above is compared to factor that gets the value of zero.

**Occlusion distance: inverse of visual demand of the driving situation.

In the model, the grand mean of the in-car glance duration is 932 milliseconds for Carrio on the larger 7” screen. Compared to the Carrio application, the use of regular smartphone applications (Android) increase the in-car glance duration by 279 milliseconds and when the size of the screen decreases from 7” to 4.5”, the duration of the in-car glance increases by 39 milliseconds. The model indicates also that one-meter increase in occlusion distance – which can be interpreted as inverse of visual demand of the driving situation – increases the duration of the in-car glance by 12 milliseconds. This means that there is a 120-milliseconds increase in in-car glance duration by 10 meter increase in occlusion distance. In other words, when the driving scenario was less visually demanding (e.g., no junctions ahead), participants were able to glance the in-car device longer.

5.2 Model 2

In order to estimate which task features affect in-car glance duration and how much, we constructed another multilevel model. The glance data from both experiments was organized in a longitudinal format and all tasks were included in the data. The data contained 14 990 in-car glances that belonged to 47 participants.

5.2.1 Task features

The dependent variable in the model was again in-car glance duration. In the intercept only model the ICC was 12.3 %. Again, this justifies the use of a multilevel model. The construction procedure of the model was identical with the previous one. Again, the inspection of the model revealed that both driving speed and age groups after NHTSA (2013) were significant in the model, affecting glance durations, but they did not significantly improve the fit of the model. In the final model (Table 10), the fixed factors were task and occlusion distance. As random factors, we had intercepts for participants. We found no significant interactions of the factors. Again, we visually inspected residual plots and they did not indicate any clear aberrations from normality or homoscedasticity.

The equation for the second model is:

$$duration_{ij} = b_0 + b_1 task_{ij} + b_2 OD_{ij} + u_{0j} + e_{0ij} \qquad (2)$$

where $duration_{ij}$ is in-car glance duration (DV), b_0 is the intercept (grand mean), $b_1 task_{ij}$ is the in-car task, $b_2 OD_{ij}$ is occlusion duration (m, inverse of visual demand of the driving situation), u_{0j} is the random effect (driver), and e_{0ij} is the residual.

Fixed effects	estimate	standard error	*p*	95 % confidence interval
Intercept	580	46	< .001	488 – 672
Email replying (Android, landscape, 4.7”, manual text entry)	747	60	< .001	627 – 867
Email reading (Android, landscape, 4.5”, manual text entry)	726	59	< .001	607 – 845
Email reading (Android, portrait, 4.5”, tapping)	693	25	< .001	645 – 741
Song searching (Android, landscape, 4.5”, manual text entry)	656	60	< .001	536 – 776
Song searching (Android, portrait, 4.5”, manual text entry)	643	25	< .001	593 – 692
Email replying (Carrio, landscape, 4.7”, speech-to-text function + tapping)	442	61	< .001	321 – 563
Song searching (Carrio, landscape, 4.5”, speech-to-text function + tapping)	429	60	< .001	309 – 549
Email reading (Carrio, landscape, 4.5”, read-aloud function + tapping)	399	61	< .001	278 – 520
Song searching (Carrio, landscape, 7”, speech-to-text function + tapping)	373	25	< .001	323 – 423
Email reading (Carrio, landscape, 7”, read-aloud function + tapping)	322	27	< .001	269 – 375
View-switching (Android, portrait, 4.5”, tapping)	303	27	< .001	251 – 356
View-switching (Carrio, landscape, 7”, swiping)	0*	0*		
Occlusion distance (m)**	11	0	< .001	9–13
Random effects	σ^2			
Intercept (participant)	35	3	< .001	
Residual	222	7	< .001	
Intraclass correlation (ICC)				
Participant	.136			
Model fit (-2RLL)	20179.88			

Table 10: Multilevel model for in-car glance duration (ms) per task – Model 2.

* The factor above is compared to factor that gets the value of zero.

**Occlusion distance: inverse of visual demand of the driving situation.

The model represents the relative visual demand of the studied tasks when controlling for the visual demands of the driving. The grand mean of the in-car glance duration is 580 milliseconds for the easiest task of view-switching with Carrio on the large screen. In Table 10, the tasks are sorted from the visually most demanding task (email replying with Android) to the visually least

demanding task (view-switching with Carrio) based on the estimates. The estimate tells how much a single in-car glance duration is expected to increase in milliseconds compared to the least demanding task.

5.2.2 Tasks grouped by visual demand

Since the tasks' estimates' 95 % confidence intervals in Model 2 (Table 10) partly overlap, reliable interpretation of the exact order of the tasks regarding their visual demand cannot be made. That is why we organized the tasks into three groups based on the overlaps in the confidence intervals: visually high demanding, visually intermediately demanding, and visually low demanding tasks (Table 11). We also added the percentages of red in-car glances to Table 11 in order to compare the visual demand (i.e., in which group the task belonged to) with the visual distraction potential of the task (i.e., red in-car glances). The tasks' red in-car glance percentages in Table 11 have a strong correlation with the visual demand estimates of the tasks in Table 10: $r = .772$ ($p = .003$, $N = 12$). In addition, we added mean number of in-car glances and mean NASA-TLX scores (Hart and Staveland, 1988) into Table 11 in order to compare these figures with the estimated visual demand and visual distraction potential of the tasks.

Visually high demanding tasks (application, screen orientation, screen size, interaction methods)	median red in-car glance % (*IQR*)	range of red in-car glance %	mean number of in-car glances (*SD*)	mean NASA-TLX (*SD*)	range of mean NASA-TLX
Email replying (Android, landscape, 4.7”, manual text entry)	12.40 (17.69)		51.42 (19.20)	51.28 (14.52)	
Email reading (Android, landscape, 4.5”, manual text entry)	13.93 (12.19)		85.63 (18.07)	51.28 (11.82)	
Email reading (Android, portrait, 4.5”, tapping)	19.00 (19.50)	8.14–19.00	86.83 (31.44)	48.68 (16.52)	48.68–56.49
Song searching (Android, landscape, 4.5”, manual text entry)	8.14 (14.08)		51.83 (15.08)	50.62 (13.88)	
Song searching (Android, portrait, 4.5”, manual text entry)	16.00 (14.25)		63.78 (23.68)	56.49 (16.38)	
Visually intermediately demanding tasks					
Email replying (Carrio, landscape, 4.7”, speech-to-text function + tapping)	4.83 (12.91)		38.50 (8.23)	38.95 (12.98)	
Song searching (Carrio, landscape, 4.5”, speech-to-text function + tapping)	8.30 (10.66)		57.21 (12.45)	39.44 (13.00)	
Email reading (Carrio, landscape, 4.5”, read-aloud function + tapping)	14.14 (16.62)	4.83–14.14	38.92 (12.33)	36.77 (16.01)	36.74–43.73
Song searching (Carrio, landscape, 7”, speech-to-text function + tapping)	6.00 (13.50)		65.14 (19.11)	43.73 (17.14)	
Email reading (Carrio, landscape, 7”, read-aloud function + tapping)	10.00 (16.00)		41.52 (12.43)	38.89 (17.65)	
View-switching (Android, portrait, 4.5”, tapping)	6.00 (8.25)		39.57 (6.81)	36.74 (15.28)	
Visually low demanding tasks					
View-switching (Carrio, landscape, 7”, swiping)	0 (3.25)	0	20.12 (9.58)	25.03 (10.50)	

Table 11: Task groups based on the confidence intervals of the multilevel model (Model 2) in Table 10.

In Table 11, all tasks labelled as visually high demanding, are tasks conducted with regular smartphone applications (Android). Three out of five tasks were performed using touch screen keyboard which in many studies has indicated high visual distraction potential (e.g., Crandall and Chaparro, 2012; McKeever et al., 2013; Reimer et al., 2014b; Tsimhoni et al., 2004). The group includes two email reading tasks and one email replying task.

All the emails in the reading tasks started with a short greeting, other than that, the emails contained on average four sentences of meaningful information. Based on the mean number of glances, participants made on average 4.34 glances per email in Experiment 1 and 4.28 glances in Experiment 2. This indicates that participants read one sentence per glance, on average. To

complete one task in the email replying task conducted with Android, participants had to read an email (1–2 meaningful short sentences) and type an answer (2–3 words, 16.75 characters on average). In addition, four button presses were required. Based on the mean number of glances, to read and type one email required on average 12.85 glances. The average should be close to 21.75 glances if the participants had typed a single character per glance. Here, end of a sentence or finishing typing a word can be assumed to be the self-selected subtask boundary that offered a natural break point to participants. In general, subtask boundaries are used while multitasking to switch attention in natural break points to reduce cognitive load (e.g., Janssen et al., 2012; Payne et al., 2007) and our findings are consistent with this idea.

The range of the red in-car glances is mainly in line with the visual demand grouping – one exception in the high visual demand group is the song searching task that had lower red in-car glance percentage than some tasks in the visually intermediately demanding group. This task was conducted with Android in landscape mode. Landscape mode slightly widens the touch screen keyboard's buttons compared to portrait mode (see Figures 4 and 7). This width difference of buttons could have caused more typing errors in portrait mode and less typing errors in landscape mode. Lee et al. (2016) found that errors during the in-car tasks, for instance, increased the duration of the in-car glances and total task time. Based on this, hypothetically, the errors in portrait mode may have caused participants to glance the in-car device at a road point where they would have not glanced without the cognitive distraction (Lee et al., 2016, 2007) caused by the typing errors. This could explain the higher red in-car glance percentage in the song searching task in the portrait mode and the lower percentage in the landscape mode. However, we were not able to measure the typing errors participants made during the tasks to test this hypothesis.

The intermediate group consisted of mainly Carrio tasks which required only button presses and either speaking (speech-to-text function) or listening (read-aloud function) but no typing. In addition, there is one Android task, view-switching, which required only two button presses to be successfully conducted. Interesting is, that in this group some of the tasks required in the software level equal amount or even more button presses to be completed (see Tables 2 and 6) than the ones in the group of visually high demanding tasks. Even though some Carrio tasks in the intermediate group had more button presses than the Android tasks in high group, together with speech-to-text and read-aloud functions these button presses may have formed subtask boundaries that were beneficial to participants for decreasing visual distraction of the tasks. Notable is, that against common belief, as discovered earlier by Reimer and Mehler (2013) and Reimer et al. (2014a), these voice-based interfaces often require visual-manual input too.

As in the group of visually high demanding tasks, the intermediate group also had one deviation among the red in-car glance percentages. Carrio's email reading task conducted with the 4.5" screen had higher red in-car glance percentage than some tasks in the visually high demanding group. The observed errors made during the selection of the next email (see Figure 8) could have an effect on these red glance percentages. That is, some participants selected same emails twice since they were confused which email was most recently selected. Alternatively, the presence of email's first line of text (see Figure 8) could have affected both red in-car glance percentages and

mean number of in-car glances: participants may have read the presented line of text instead of just listening to the email. Again, hypothetically, the errors may have caused participants to glance the in-car device at a road point where they would have not glanced without this cognitive distraction (Lee et al., 2016, 2007). However, unfortunately we do not have access to error data in these tasks. Otherwise, the range of the red in-car glance percentages is well in line with the visual demand grouping.

Finally, there was only one task in the group of visually low demanding tasks: view-switching task with Carrio. This task was relatively easy since the order of the views seemed to be easily learnable and this enabled drivers to switch views with a simple gesture while looking at the road ahead. This task had the lowest mean number of in-car glances (see Table 3). This is a similar finding as in previous studies that found simple gestures for scrolling pages one-by-one to be the most visually least demanding and distracting when compared to button presses or kinetic scrolling (e.g., Kujala, 2013; Lasch and Kujala, 2012).

Additionally, the range of mean NASA-TLX scores (Hart and Staveland, 1988) – which measure subjective task workload – is well in line with the visual demand grouping of the tasks. This suggests that subjectively experienced task workload is particularly connected with the visual demand of the tasks.

6 GENERAL DISCUSSION

We conducted two driving simulator experiments with 48 participants in order to study the impacts of touch screen size, interaction methods, and subtask boundaries on secondary task's visual demand and visual distraction potential. For controlling the visual demand of the driving situation and participants' individual differences in in-car glance durations, we utilized multilevel modeling.

In Experiment 1, automotive-targeted application (Carrio) was running in a 7" tablet in landscape mode and was compared to regular smartphone (4.5", portrait mode) applications (Android). The distraction potential of the tested tasks was assessed with a novel method introduced by Kujala and Mäkelä (2015) which categorizes part of the in-car glances into red in-car glances, that is, inappropriately long in-car glances in relation to the visual demand of the given driving situation (i.e., visual distraction). This novel testing method allowed us to compare visual distraction potential of the tested tasks when the visual demands of the driving scenario was controlled for. In Experiment 1 Carrio had significantly lower percentages of red in-car glances in each task compared to the tasks conducted with regular smartphone applications (Research Question 1). Since there were two confounding factors, screen size and orientation of the device, we could not exactly point out which design factors caused Carrio's lower red in-car percentages, and therefore we conducted another experiment. In Experiment 2, both Carrio and Android tasks were conducted with a smartphone in landscape mode. Based on the distraction potential testing, only Carrio's email replying task had significantly lower red in-car glance percentage than any Android task (Research Question 1).

Since these results from Experiment 1 and Experiment 2 individually did not clarify the effects of the different design factors, we constructed two multilevel models based on the data from both

experiments. With multilevel models we investigated how screen size, screen orientation, application, and task affect in-car glance durations in the tested tasks. Multilevel modeling enabled us to control the effects of visual demand of the driving scenario and individual differences on in-car glance durations. Together with the accompanying data from the experiments, these models indicated that, in general, Carrio tasks had lower visual demand and visual distraction potential compared to the tasks conducted with the regular smartphone applications (Research Questions 1 and 2).

It is intuitive to think that bigger touch screen size enables more efficient task performance (e.g., Hancock et al., 2015; Raptis et al., 2013). Based on the multilevel models, the 2.5 inches larger touch screen slightly diminished the durations (39 ms) of in-car glances (see Model 1) as well as visual demand and visual distraction potential of the secondary task (see Model 2 and Table 11) (Research Question 3). However, the effect was surprisingly small. To our best knowledge, this was the first controlled study that investigated glance durations in the automotive context regarding the effects of touch screen size.

The application had larger relative impact than the screen size – use of the automotive-targeted Carrio application decreased the duration of in-car glances by 279 milliseconds (Model 1) as well as the secondary tasks' visual demand and visual distraction potential (see Model 2 and Table 11) compared to regular smartphone applications (Research Questions 1 and 2). This implies that the application's interaction methods may be more crucial than the size of the in-vehicle screen used for safe use while driving (Research Question 3). There was no effect of orientation of the device on in-car glance durations (Research Question 3), which was discovered also in the study of Lasch and Kujala (2012). Model 1 (Table 9) also indicated that when the driving scenario was less visually demanding (e.g., no junctions ahead), participants were able to glance the in-car device longer. This finding is consistent with, for instance, Wierwille's, (1993) visual sampling model and endorses Kircher and Ahlström's (2017) proposal about the minimum required attention for each driving scenario, which can be achieved with diverse patterns of visual sampling.

Further, based on the overlaps in the confidence intervals in the Model 2, we identified three task groups (see Table 11): visually high demanding, visually intermediately demanding, and visually low demanding tasks (Research Question 2) which all have their own common features (Research Question 3). The main features of visually high demanding tasks were touch screen typing and self-selected subtask boundaries – which are not derived from the user interface. All of these tasks were also conducted with the regular smartphone applications (Android). The main feature of the visually intermediate task group was the invocation of speech-to-text and read-aloud functions as well as the automotive-targeted application design. Because of the design, all of the visual-manual interactions could be easily split into brief visual encoding – single button press steps without inducing cognitive load for keeping in mind the task state during on-road glancing. Finally, the visually least demanding task group contained only one task which required only simple swiping gestures at any point of the touch screen and visual confirmation of the target view. The measured red in-car glance percentages and experienced task workload were generally well in line with the visual demand grouping.

Additionally, we found a plausible impact of subtask boundaries on the visual demand and distraction potential of the tested tasks. As mentioned above, one common feature of the visually high demanding tasks were self-selected subtask boundaries. Similarly, one common feature of visually intermediately demanding tasks was that all the visual-manual interactions could be effortlessly split into small subtasks of button presses. Together with speech-to-text and read-aloud functions these button presses may have formed beneficial subtask boundaries for participants reducing visual demand and distraction potential of the tasks. Based on these findings, increase in the preferred number of visual or visual-manual interaction steps during an in-car glance (e.g., pressing one button vs. typing one word), increases both the duration of the in-car glance as well as its visual distraction potential. These observations of subtask boundaries support the previous findings of, for instance, Janssen et al. (2012), Lee et al. (2015) and Lee and Lee (2019).

Interesting discovery was that some Carrio tasks had higher mean number of glances per task than the corresponding Android tasks. Regardless of that, Carrio's in-car glance duration estimates and red in-car glance percentages were lower than or at the same level with Android in these tasks. This indicates that the mean number of glances is not alone a sufficient metric for assessing in-car task's visual demand or visual distraction. Therefore, we suggest that visual demand of the tasks is not necessarily equal to visual distraction caused by the tasks. For instance, NHTSA's driver distraction guidelines for in-vehicle electronic devices (2013) are based on static glance metrics which are supposed to determine if a certain task is visually distracting or not. NHTSA's (2013) guidelines seem to measure, before anything, visual demands of the tasks, not visual distraction per se. Based on this study, glance metrics – at least alone – cannot specify if a task is distracting or not since the visual demands of the driving situation have an impact on the glance durations, and even more importantly, on how distractive the particular in-car glance is. Besides testing and regulation of in-vehicle devices, this is important to be realized in the development of risk-based insurance systems (e.g., Yin and Chen, 2018) , and distraction warning and other driver monitoring systems (e.g., Hu et al., 2017; Wu et al., 2013; Yin et al., 2018).

6.1 Limitations and further research

The presented results concern only the type of tasks that we studied in this paper. To analyze even more carefully the user interface design factors that could diminish visual distraction, other types of tasks should be studied and preferably in experimental designs with lower number of variables. The level and generalizability of the analysis could be further improved by extracting the glances related to particular features of the task. However, here our general aim was to investigate, if, and to what extent, an automotive-targeted application can reduce visual distraction potential of real-world in-car tasks (without splitting these into subtasks) compared to regular smartphone applications.

Unfortunately, we were not monitoring the typing errors participants possibly made. Task errors and the associated recoveries have an effect on the number of the in-car glances as well as glance durations (e.g., Lee et al., 2016) and therefore are something that should be taken into consideration in the analyses in the future.

Another important point of view is the acceptance of the interaction methods used in automotive-targeted applications. It should be further studied which kind of interaction methods – that should diminish drivers' visual inattention – drivers accept and are willing to use during driving. Whereas read-aloud function could decrease visual distraction, it does not achieve this in real life, if drivers prefer to read the messages. For instance, they may find listening messages too slow compared to reading them. Since Carrio's read-aloud function – with the possibility to read the first lines of the messages – produced high number of in-car glances, this could suggest that read-aloud function is not the most accepted interaction modality for all kind of in-car tasks.

7 CONCLUSIONS

Despite legislation, people are still talking on the phone, dialing, texting (Oviedo-Trespalacios et al., 2016), and even playing games (Ahlström et al., 2019; Mäkelä and Kujala, 2017) while driving. User interfaces that are designed for the automotive context and accepted by the drivers could be a solution to diminish visual distraction by smartphones. In order to be able to design automotive-targeted user interfaces, the design factors' visual distraction potential should be better understood.

In this paper, we conducted two driving simulator experiments with 48 participants in order to study the effects of touch screen size, user interface design, and subtask boundaries on secondary task's visual demand and visual distraction potential. With multilevel modeling, we controlled the effects of visual demand of the driving scenario and individual differences on in-car glance durations. To our best knowledge, this was the first study that investigated the selected application features' effects and screen size on in-car tasks' visual demand and distraction potential while controlling for the varying visual demand of the driving situation.

The findings indicate the potential of well-designed and driver-friendly in-car user interfaces to decrease visual demands of in-car tasks and the associated visual distraction potential compared to use of regular smartphone applications. In addition, a small impact of 2.5" larger in-vehicle screen size decreasing in-car glance durations and diminishing visual demand and visual distraction potential of the secondary task was found. However, the effect of screen size was small. In line with previous research (e.g., Janssen et al., 2012; Lee and Lee, 2019), drivers' ability to break down an in-car task into smaller subtasks (e.g., pressing one button vs. typing one word) seem to decrease in-car glance durations and enable better adjustment of glancing behavior in relation to the demands of the driving situation.

The most important methodical discovery in the present study comes from the dissociation of the visual demand from the visual distraction potential in two of the tasks. Also, some of the in-car tasks required high number of in-car glances even though the measured visual distraction potential (i.e., red in-car glance percentage) was low. Even if increasing visual demand of a task – as measured by in-car glance duration or number of glances – may increase its visual distraction potential, these two are not necessarily equal. Another notable observation was that when the visual demand of the driving situation decreased, the durations of the in-car glances increased. This finding is in line with, for instance, Wierwille's, (1993) visual sampling model and supports the suggestion of Kircher and Ahlstrom (2017) about the minimum required attention for each driving

scenario that can be fulfilled with various patterns of visual sampling. Therefore, a red in-car glance can be interpreted as a failure to reach the minimum required attention in the particular driving situation – or, in other words, visual distraction.

ACKNOWLEDGEMENTS

This work was partly funded by Ficonic Solutions Ltd. The funder had no role in data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

Ahlström, C., Wachtmeister, J., Nyman, M., Nordenström, A., Kircher, K., 2019. Using smartphone logging to gain insight about phone use in traffic. Cogn. Technol. Work 1–11. https://doi.org/10.1007/s10111-019-00547-6

Bayer, J.B., Campbell, S.W., 2012. Texting while driving on automatic: Considering the frequency-independent side of habit. Comput. Human Behav. 28, 2083–2090. https://doi.org/10.1016/J.CHB.2012.06.012

Beckers, N., Schreiner, S., Bertrand, P., Mehler, B., Reimer, B., 2017. Comparing the demands of destination entry using Google Glass and the Samsung Galaxy S4 during simulated driving. Appl. Ergon. 58, 25–34. https://doi.org/10.1016/j.apergo.2016.05.005

Broström, R., Aust, M.L., Wahlberg, L., Källgren, L., 2013. What drives off-road glance durations during multitasking – capacity, practice or strategy? Proc. 3rd Int. Conf. Driv. Distraction Ina.

Broström, R., Bengtsson, P., Aust, M.L., 2016. Individual glance strategies and their effect on the NHTSA visual manual distraction test. Transp. Res. Part F Traffic Psychol. Behav. 36, 83–91. https://doi.org/10.1016/J.TRF.2015.10.017

Caird, J.K., Johnston, K.A., Willness, C.R., Asbridge, M., Steel, P., 2014. A meta-analysis of the effects of texting on driving. Accid. Anal. Prev. 71, 311–318. https://doi.org/10.1016/j.aap.2014.06.005

Choudhary, P., Velaga, N.R., 2017. Modelling driver distraction effects due to mobile phone use on reaction time. Transp. Res. Part C Emerg. Technol. 77, 351–365. https://doi.org/10.1016/j.trc.2017.02.007

Crandall, J.M., Chaparro, A., 2012. Driver distraction: Effects of text entry methods on driving performance, in: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. pp. 1693–1697. https://doi.org/10.1177/1071181312561339

Gaffar, A., Kouchak, S.M., 2017. Quantitative driving safety assessment using interaction design benchmarking, in: 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). IEEE, San Francisco, pp. 1–8. https://doi.org/10.1109/UIC-ATC.2017.8397626

Gauld, C.S., Lewis, I., White, K.M., Fleiter, J.J., Watson, B., 2017. Smartphone use while driving: What factors predict young drivers' intentions to initiate, read, and respond to social interactive technology? Comput. Human Behav. 76, 174–183. https://doi.org/10.1016/J.CHB.2017.07.023

Grahn, H., Kujala, T., 2018. Visual distraction effects between in-vehicle tasks with a smartphone and a motorcycle helmet-mounted head-up display, in: ACM International Conference

Proceeding Series. pp. 153–162. https://doi.org/10.1145/3275116.3275134

Guo, F., Klauer, S.G., Hankey, J.M., Dingus, T.A., 2010. Near crashes as crash surrogate for naturalistic driving studies. Transp. Res. Rec. J. Transp. Res. Board 2147, 66–74. https://doi.org/10.3141/2147-09

Hancock, P.A., Sawyer, B.D., Stafford, S., 2015. The effects of display size on performance. Ergonomics 58, 337–354. https://doi.org/10.1080/00140139.2014.973914

Hart, S.G., 2006. Nasa-Task Load Index (NASA-TLX); 20 years later. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 904–908. https://doi.org/10.1177/154193120605000909

Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Adv. Psychol. 52, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

He, J., Chaparro, A., Nguyen, B., Burge, R.J., Crandall, J., Chaparro, B., Ni, R., Cao, S., 2014. Texting while driving: Is speech-based text entry less risky than handheld text entry? Accid. Anal. Prev. 72, 287–295. https://doi.org/10.1016/j.aap.2014.07.014

He, J., Chaparro, A., Wu, X., Crandall, J., Ellis, J., 2015a. Mutual interferences of driving and texting performance. Comput. Human Behav. 52, 115–123. https://doi.org/10.1016/J.CHB.2015.05.004

He, J., Choi, W., McCarley, J.S., Chaparro, B.S., Wang, C., 2015b. Texting while driving using Google Glass™: Promising but not distraction-free. Accid. Anal. Prev. 81, 218–229. https://doi.org/10.1016/j.aap.2015.03.033

Hox, J., 1998. Multilevel modeling: When and why, in: Balderjahn, I., Mathar, R., Schader, M. (Eds.), Classification, Data Analysis, and Data Highways. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, pp. 147–154. https://doi.org/https://doi.org/10.1007/978-3-642-72087-1_17

Hu, J., Xu, L., He, X., Meng, W., 2017. Abnormal driving detection based on normalized driving behavior. IEEE Trans. Veh. Technol. 66, 6645–6652. https://doi.org/10.1109/TVT.2017.2660497

Jamson, A.H., Westerman, S.J., Hockey, G.R.J., Carsten, O.M.J., 2004. Speech-based e-mail and driver behavior: Effects of an in-vehicle message system interface. Hum. Factors 46, 625–639. https://doi.org/10.1518/hfes.46.4.625.56814

Janssen, C.P., Brumby, D.P., Garnett, R., 2012. Natural break points: The influence of priorities and cognitive and motor cues on dual-task interleaving. J. Cogn. Eng. Decis. Mak. 6, 5–29. https://doi.org/10.1177/1555343411432339

Kircher, K., Ahlstrom, C., 2017. Minimum required attention: A human-centered approach to driver inattention. Hum. Factors 59, 471–484. https://doi.org/10.1177/0018720816672756

Kujala, T., 2013. Browsing the information highway while driving: Three in-vehicle touch screen scrolling methods and driver distraction, in: Personal and Ubiquitous Computing. https://doi.org/10.1007/s00779-012-0517-2

Kujala, T., Grahn, H., 2017. Visual distraction effects of in-car text entry methods: Comparing keyboard, handwriting and voice recognition, in: Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. pp. 1–10. https://doi.org/10.1145/3122986.3122987

Kujala, T., Grahn, H., Mäkelä, J., Lasch, A., 2016a. On the visual distraction effects of audio-visual route guidance, in: Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. pp. 169–176. https://doi.org/10.1145/3003715.3005421

Kujala, T., Lasch, A., Mäkelä, J., 2014. Critical analysis on the NHTSA acceptance criteria for in-vehicle electronic devices, in: Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. pp. 1–8. https://doi.org/10.1145/2667317.2667322

Kujala, T., Mäkelä, J., 2018. Naturalistic study on the usage of smartphone applications among Finnish drivers. Accid. Anal. Prev. 115, 53–61. https://doi.org/10.1016/j.aap.2018.03.011

Kujala, T., Mäkelä, J., 2015. Testing environment and verification procedure for in-car tasks with dynamic self-paced driving scenarios, in: Proceedings of the 4th International Conference on Driver Distraction and Inattention.

Kujala, T., Mäkelä, J., Kotilainen, I., Tokkonen, T., 2016b. The attentional demand of automobile driving revisited: Occlusion distance as a function of task-relevant event density in realistic driving scenarios. Hum. Factors 58, 163–180. https://doi.org/10.1177/0018720815595901

Kujala, T., Salvucci, D.D., 2015. Modeling visual sampling on in-car displays: The challenge of predicting safety-critical lapses of control. Int. J. Hum. Comput. Stud. 79, 66–78. https://doi.org/10.1016/j.ijhcs.2015.02.009

Kujala, T., Silvennoinen, J., Lasch, A., 2013. Visual-manual in-car tasks decomposed: text entry and kinetic scrolling as the main sources of visual distraction, in: Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. pp. 82–89. https://doi.org/10.1145/2516540.2516562

Lasch, A., Kujala, T., 2012. Designing browsing for in-car music player: Effects of touch screen scrolling techniques, items per page and screen orientation on driver distraction, in: Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. pp. 41–48. https://doi.org/10.1145/2390256.2390262

Lee, J.D., Caven, B., Haake, S., Brown, T.L., 2001. Speech-based interaction with in-vehicle computers: The effect of speech-based e-mail on drivers' attention to the roadway. Hum. Factors 43, 631–640. https://doi.org/10.1518/001872001775870340

Lee, J.D., Roberts, S.C., Hoffman, J.D., Angell, L.S., 2012. Scrolling and driving: How an MP3 player and its aftermarket controller affect driving performance and visual behavior. Hum. Factors 54, 250–263. https://doi.org/10.1177/0018720811429562

Lee, J.Y., Gibson, M., Lee, J.D., 2015. Secondary task boundaries influence drivers' glance durations, in: Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. pp. 273–280. https://doi.org/10.1145/2799250.2799269

Lee, J.Y., Gibson, M.C., Lee, J.D., 2016. Error recovery in multitasking while driving, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. pp. 5104–5113. https://doi.org/10.1145/2858036.2858238

Lee, J.Y., Lee, J.D., 2019. Modeling microstructure of drivers' task switching behavior. Int. J. Hum. Comput. Stud. 125, 104–117. https://doi.org/10.1016/J.IJHCS.2018.12.007

Lee, J.Y., Lee, J.D., 2017. Multi-level analysis of distracted drivers' glances: Enhancing the robustness of the NHTSA acceptance criteria, in: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. pp. 1919–1923. https://doi.org/10.1177/1541931213601960

Lee, Y.-C., Lee, J.D., Ng Boyle, L., 2007. Visual attention in driving: The effects of cognitive load and visual disruption. Hum. Factors J. Hum. Factors Ergon. Soc. 49, 721–733. https://doi.org/10.1518/001872007x215791

Mäkelä, J., Kujala, T., 2017. Naturalistic driving study on the usage of smart phone applications while driving, in: Proceedings of the 5th International Conference on Driver Distraction and Inattention. pp. 1–18.

McKeever, J.D., Schultheis, M.T., Padmanaban, V., Blasco, A., 2013. Driver performance while texting: Even a little is too much. Traffic Inj. Prev. 14, 132–137. https://doi.org/10.1080/15389588.2012.699695

National Highway Traffic Safety Administration, 2013. Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices, Docket No. NHTSA-2010-0053. https://doi.org/NHTSA-2010-0053

Oviedo-Trespalacios, O., Haque, M.M., King, M., Washington, S., 2016. Understanding the impacts of mobile phone distraction on driving performance: A systematic review. Transp. Res. Part C Emerg. Technol. 72, 360–380. https://doi.org/10.1016/J.TRC.2016.10.006

Owens, J.M., McLaughlin, S.B., Sudweeks, J., 2011. Driver performance while text messaging using handheld and in-vehicle systems. Accid. Anal. Prev. 43, 939–947. https://doi.org/10.1016/j.aap.2010.11.019

Payne, S.J., Duggan, G.B., Neth, H., 2007. Discretionary task interleaving: Heuristics for time allocation in cognitive foraging. J. Exp. Psychol. Gen. 136. https://doi.org/10.1037/0096-3445.136.3.370

Raptis, D., Tselios, N., Kjeldskov, J., Skov, M.B., 2013. Does size matter? Investigating the impact of mobile phone screen size on users' perceived, effectiveness and efficiency. Proc. 15th Int. Conf. Human-computer Interact. with Mob. devices Serv. - MobileHCI '13 127–136. https://doi.org/10.1145/2493190.2493204

Reimer, B., Mehler, B., 2013. The effects of a production level "voice-command" interface on driver behavior: Summary findings on reported workload, physiology, visual attention, and driving performance, MIT AgeLab Technical Report No. 2013-17A. Massachusetts Institute of Technology, Cambridge, MA. https://doi.org/2013-18A

Reimer, B., Mehler, B., Dobres, J., McAnulty, H., Mehler, A., Munger, D., Rumpold, A., 2014a. Effects of an "expert mode" voice command system on task performance, glance behavior & driver physiology, in: Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. pp. 1–9. https://doi.org/10.1145/2667317.2667320

Reimer, B., Mehler, B., Donmez, B., 2014b. A study of young adults examining phone dialing while driving using a touchscreen vs. a button style flip-phone. Transp. Res. Part F Traffic Psychol. Behav. 23, 57–68. https://doi.org/10.1016/j.trf.2013.12.017

Richter, T., 2006. What is wrong with ANOVA and multiple regression? Analyzing sentence reading

times with hierarchical linear models. Discourse Process. 41, 221–250. https://doi.org/10.1207/s15326950dp4103_1

Rumschlag, G., Palumbo, T., Martin, A., Head, D., George, R., Commissaris, R.L., 2015. The effects of texting on driving performance in a driving simulator: The influence of driver age. Accid. Anal. Prev. 74, 145–149. https://doi.org/10.1016/j.aap.2014.10.009

Salvucci, D.D., Kujala, T., 2016. Balancing structural and temporal constraints in multitasking contexts, in: Proc. of CogSci 2016.

Senders, J.W., Kirstofferson, A.B., Levison, W.H., Dietrich, C.W., Ward, J.L., 1967. The attentional demand of automobile driving. Highw. Res. Rec.

Society of Automotive Engineers, 2000. Experimental measures related to the specification of driver visual behavior using video based techniques. SAE-J2396.

Tivesten, E., Dozza, M., 2015. Driving context influences drivers' decision to engage in visual–manual phone tasks: Evidence from a naturalistic driving study. J. Safety Res. 53, 87–96. https://doi.org/10.1016/J.JSR.2015.03.010

Tsimhoni, O., Smith, D., Green, P., 2004. Address entry while driving: Speech recognition versus a touch-screen keyboard. Hum. Factors 46, 600–610. https://doi.org/10.1518/hfes.46.4.600.56813

Wierwille, W.W., 1993. An initial model of visual sampling of in-car displays and controls., in: A.G. Gale, I. D. Brown, C. M. Haslegrave, H. W. Kruysse, and S. P. Taylor, (Eds.), Vision in Vehicles IV.

Wu, B.F., Chen, Y.H., Yeh, C.H., Li, Y.F., 2013. Reasoning-based framework for driving safety monitoring using driving event recognition. IEEE Trans. Intell. Transp. Syst. 14, 1231–1241. https://doi.org/10.1109/TITS.2013.2257759

Yin, J.-L., Chen, B.-H., 2018. An advanced driver risk measurement system for usage-based insurance on big driving data. IEEE Trans. Intell. Veh. 3, 585–594. https://doi.org/10.1109/tiv.2018.2874530

Yin, J.L., Chen, B.H., Lai, K.H.R., Li, Y., 2018. Automatic dangerous driving intensity analysis for advanced driver assistance systems from multimodal driving signals. IEEE Sens. J. 18, 4785–4794. https://doi.org/10.1109/JSEN.2017.2765315