Transportation Research Part F 79 (2021) 23–34

Contents lists available at ScienceDirect

Transportation Research Part F

journal homepage: www.elsevier.com/locate/trf

Refining distraction potential testing guidelines by considering differences in glancing behavior

Hilkka Grahn *, Toni Taipalus

University of Jyväskylä, P.O. Box 35, FI-40014, Finland

ARTICLE INFO

Article history:
Received 23 September 2020
Received in revised form 9 March 2021
Accepted 16 March 2021

Keywords:
Driver inattention
Driver distraction
Visual distraction
Occlusion distance
Distraction potential testing
Individual differences

ABSTRACT

Driver distraction is a recognized cause of traffic accidents. Although the well-known guidelines for measuring distraction of secondary in-car tasks were published by the United States National Highway Traffic Safety Administration (NHTSA) in 2013, studies have raised concerns on the accuracy of the method defined in the guidelines, namely criticizing them for basing the diversity of the driver sample on driver age, and for inconsistent between-group results. In fact, it was recently discovered that the NHTSA driving simulator test is susceptible to rather fortuitous results when the participant sample is randomized. This suggests that the results of said test are highly dependent on the selected participants, rather than on the phenomenon being studied, for example, the effects of touch screen size on driver distraction. As an attempt to refine the current guidelines, we set out to study whether a previously proposed new testing method is as susceptible to the effects of participant randomization as the NHTSA method. This new testing method differs from the NHTSA method by two major accounts. First, the new method considers occlusion distance (i.e., how far a driver can drive with their vision covered) rather than age, and second, the new method considers driving in a more complex, and arguably, a more realistic environment than proposed in the NHTSA guidelines. Our results imply that the new method is less susceptible to sample randomization, and that occlusion distance appears a more robust criterion for driver sampling than merely driver age. Our results are applicable in further developing driver distraction guidelines and provide empirical evidence on the effect of individual differences in drivers' glancing behavior.

1. Introduction

Driver inattention – which is often caused by digital devices used during driving – is a universally recognized phenomenon that is connected to accidents and near-accidents in traffic (e.g., Bayer & Campbell, 2012; Caird et al., 2014; Choudhary & Velaga, 2017; Gauld et al., 2017; Guo et al., 2010; He et al., 2015; Oviedo-Trespalacios et al., 2016; Rumschlag et al., 2015; Tivesten & Dozza, 2015). Consequently, there is a large body of literature concerning distraction potential of different in-car tasks and interaction methods (i.e., how distractive these are for drivers) (e.g., Buchhop et al., 2017; Crandall & Chaparro, 2012; He, Chaparro, et al., 2015; He, Choi, et al., 2015; Kujala & Grahn, 2017; Lasch & Kujala, 2012; Ng & Brewster, 2017; Perlman et al., 2019; Reimer & Mehler, 2013; Villalobos-zúñiga et al., 2016). Although driver inattention has received ample scholarly attention, there is no commonly agreed definition for driver inattention or driver

* Corresponding author.
E-mail addresses: hilkka.grahn@jyu.fi (H. Grahn), toni.taipalus@jyu.fi (T. Taipalus).

https://doi.org/10.1016/j.trf.2021.03.009
1369-8478/

distraction, which in turn affects the ways those are operationalized and measured. Regan et al. (2011) have proposed a taxonomy regarding driver inattention. According to the taxonomy, driver distraction is a form of driver inattention, and being distracted requires some competing activity while driving. The taxonomy also suggests that a driver can be inattentive while not being distracted, but not be distracted without being inattentive. This categorization (Regan et al., 2011) of driver distraction being a subcategory of driver inattention is adopted here.

To answer the problem of driver distraction caused by digital devices, the National Highway Traffic Safety Administration (NHTSA), published guidelines (Driver Distraction Guidelines for In-Vehicle Electronic Devices) in 2013 for measuring and assessing how distractive different in-car tasks are. In the method (NHTSA, 2013), distraction testing is conducted in a driving simulator while driving 50 miles per hour on a straight 4-lane road, following a lead vehicle. Three metrics, measured using eye-tracking technology, are used to evaluate the in-car task at hand: total glance time, mean glance duration, and the percentage of over 2-second glances. In more detail, these metrics mean that *1)* the total glance time should not exceed 12 s when performing a task; *2)* the mean glance time should be less than or equal to 2 s when performing a task; and *3)* the percentage of over 2-second glances should not exceed 15% of the total number of in-car glances. According to the guidelines, testing should be conducted with 24 randomly selected participants who are divided into four groups of six, according to participant age (18–24 years, 25–39 years, 40–54 years old, and older than 55 years). Although the NHTSA guidelines have provided the field of distraction testing with a solid starting point, the guidelines have received many suggestions for improvements, especially regarding individual glancing behaviors and the visual demands of the driving scenario.

Broström et al. (2013) studied participants' glance durations while driving and conducting secondary tasks and noticed that participants who exceeded the 2-second glance duration limit were often the same participants. These participants were labeled as *long glancers*. Donmez et al. (2010) studied young drivers and were able to identify within one driver sample three groups of drivers based on their glancing behavior: low-risk drivers, moderate-risk drivers, and high-risk drivers. Similarly, Broström et al. (2016) identified four individual glancing strategies among their participants in a driving simulator study: *optimizers*, *normal glancers*, *long glancers*, and *frequent glancers*. In addition, these long glancers and frequent glancers were participants whose glancing strategy affected the reliability of the results of testing done according to the NHTSA (2013) guidelines. These studies indicate that drivers have individual in-car glance durations, that is, individual glancing behavior, which seems to be a relatively constant individual feature.

Additionally, Kujala et al. (2014) examined the NHTSA's acceptance criteria for in-car tasks in a driving simulator study and observed that participants had individual differences regarding how they experienced levels of visual demand. Levels of visual demand were measured utilizing occlusion times, that is, how long participants were willing to drive without visual information. This observation means, in addition to individual in-car glance durations, that drivers also have individual differences regarding how long they prefer to drive without visual information, when the task is to concentrate on safe driving.

As mentioned before, in the NHTSA testing method, a straight 4-lane road is used as a driving scenario. According to, for instance, Tivesten and Dozza (2014), Tsimhoni and Green (2001) as well as Wierwille (1993), the visual demands of the driving scenario have an effect on off-road glance durations. Therefore, it has been criticized that NHTSA's method does not account either the visual demands of the driving scenario, or its effect on glancing behavior (e.g., Kujala et al., 2014; Tivesten & Dozza, 2015), both of which have an effect on driver inattention.

Since drivers seem to have individual variation in glancing behavior and the visual demands of the driving scenario also affect glance durations, it would be logical to take these issues into consideration when testing the distraction potential of in-car tasks. In point of fact, studies by Broström et al. (2013, 2016), Kujala et al. (2014), J. Y. Lee and Lee (2017), and Ljung Aust et al. (2015) indicated the need for developing a more robust distraction potential testing method that would consider individual glancing behavior. In addition, Kircher et al. (2019) suggest reconsidering these fixed glance durations as indicators of distraction. Furthermore, Broström et al. (2016), and J. Y. Lee and Lee, (2017) tested the effects of individual glancing behavior on the results of the distraction potential testing conducted following the NHTSA guidelines. They noticed that neglecting these individual factors can lead to a situation in which the results of the distraction potential testing are highly dependent on the driver sample, and not on the phenomenon studied. Furthermore, using data from a test conforming to the NHTSA guidelines, Ljung Aust et al. (2015) randomized 50 test groups of 24 drivers from a participant pool of 48 participants, and discovered that the distraction potential test results had "near stochastic outcomes".

Several solutions to account for the individual differences in glancing behavior when conducting distraction potential testing have been tested, for instance, Intolerance of Uncertainty (Kujala, Grahn, et al., 2016), visual short term working memory (Kujala & Grahn, 2017), and individual performance capacity measured with Trail Making Test (Broström et al., 2013), yet none of these measures have been shown to have an association with occlusion distance (OD, i.e., how far a driver can drive with their vision covered) or glancing behavior. One solution is to take the ages of participants into consideration, and it has been noted that age is one factor affecting glance durations: the higher the age, the longer the glance duration (e.g., Dobres et al., 2016; J. Lee et al., 2015; Son & Park, 2012; Wikman & Summala, 2005). In the NHTSA guidelines, participants in one group should be older than 55 years. As said, this age grouping could be one way to consider individual differences in glancing behavior. According to Domeyer et al. (2014), this oldest age group is most likely to cause the tested task to fail the distraction potential testing, that is, the secondary task is considered distracting. This, however, propounds the question of whether the included oldest drivers are long glancers, as it has been shown that some younger drivers are also long glancers, and not all older drivers are necessary long glancers. If the purpose is to obtain a diverse sample, the criterion for diversity should be based on *driver differences*, not factors that have been statistically shown to affect said differences.

Inspired by the study by Ljung Aust et al. (2015), where the authors observed that the randomization of the participant sample affects the results obtained by using the NHTSA (2013) method, we set out to study whether a *new*, occlusion distance-based method proposed by Kujala and Mäkelä (2015) is susceptible to similar participant randomization. While we explain the new method in more detail in Section 2.1, it is worth noting that the method differs from the NHTSA guidelines (2013) by considering occlusion distance rather than participant age, and that the driving scenario involves a suburban environment with turns and intersections, rather than a straight road. To study this phenomenon, we sampled 23 participants from a pool of 46 participants who used Android applications in two driving simulator experiments (Grahn & Kujala, 2020), and these two experiments utilized the new distraction potential testing method (i.e., the method described in Kujala & Mäkelä, 2015). The hypotheses tested in this study are:

H1. *The results of the new occlusion distance-based distraction potential testing method do not change when the sample of participants is randomized.*

H2. *The results of the new occlusion distance-based distraction potential testing method change when only participants with low occlusion distance* ($Mdn \leq 16$ *m*) *are selected.*

In other words, the first hypothesis inspects the robustness of driver sampling based on occlusion distance. The results are compared to those reported in the study of Grahn and Kujala (2020), who utilized the new method instead of that proposed in the NHTSA guidelines (2013). The second hypothesis is concerned with validating the results of the first hypothesis with this dataset, and with indicating whether occlusion distance can be used as a validation criterion. In other words, the second hypothesis tests if the participants in the dataset can be selected in a way that *may affect the results to begin with*. Furthermore, if the participants are handpicked based on occlusion distances, and this affects the results, it indicates that occlusion distance is indeed related to glancing behavior. Finally, *change* mentioned in the hypotheses relates to the change whether a task is deemed distracting or not by the distraction potential test.

2. Materials and methods

2.1. The new distraction testing method

In order to account for drivers' individual differences in glancing behavior, we used a distraction testing method introduced by Kujala and Mäkelä (2015) in the experiments. The method in question has been previously used in studies by Grahn and Kujala (2020, 2018), Kujala, Grahn, et al. (2016) as well as Kujala and Grahn (2017), and is based on a study by Kujala, Mäkelä, et al. (2016). This new distraction testing method utilizes a visual occlusion technique which was initially introduced by Senders et al. (1967). The original purpose of the occlusion technique was to investigate aspects of visual information processing performance while driving (Milgram, 1987). According to the occlusion technique, the driver is instructed to maintain safe driving, while striving to keep their vision occluded as much as possible (Senders et al., 1967). In the original occlusion technique by Senders et al. (1967), driver's vision was occluded (i.e., driving blind), and the *time* driven with occluded vision was measured. Contrary to the original technique, this new testing method measured occlusion *distance*, not occlusion time. Occlusion distance refers to a driver's preferred distance in meters that is driven during the occluded period. Measuring occlusion distance allows drivers to freely adjust their driving speed if needed. Arguably, as drivers have different qualities, the distance of occlusions is dependent on the driver – some drivers accept uncertainty induced by the lack of vision more effectively than others (Kujala, Mäkelä, et al., 2016).

The new testing method is based on a study by Kujala, Mäkelä, et al. (2016) where participants' ($N = 97$) occlusion distances on simulated highway and suburban roads were measured and later mapped to the test routes. We refer to the results of this process as the *occlusion distance map* (ODM). In the map (Fig. 1), every 1 X 1 m route point contains information on the occlusion distance that was driven in that route point. The occlusion distance map serves two purposes. First, the routes are used for driver sample validation to ensure that the driver sample matches the occlusion distance distribution of the original driver sample of 97 drivers (Kujala, Mäkelä, et al., 2016), and contains drivers with different glancing behaviors – from those who are able to drive longer occlusion distances to those who are able to drive shorter occlusion distances. As discussed in the Introduction, the NHTSA method (2013) utilizes age groups to facilitate a diverse driver sample, and although it has been shown that age is related to glance durations (e.g., Dobres et al., 2016), not all older drivers are long glancers, and not all younger drivers are short glancers. This in turn indicates that age as a validation criterion should be scrutinized. Second, the occlusion distance map separates the relatively demanding route points from the relatively undemanding, with the indication that different parts of the route are more demanding than others, and this needs to be accounted in any subsequent analyses.

Suburban roads are used for the actual distraction testing. During the distraction testing, the in-car glances (i.e., glances directed towards an in-vehicle device) are categorized as appropriate or inappropriate glances based on the distance driven during an in-car glance from a particular route point where the in-car glance begins. These inappropriate glances – or *red in-car glances* – refer to an in-car glance length that exceeds the 85th percentile of the original experiment's driver sample ($N = 97$, Kujala, Mäkelä, et al., 2016) on that route point. These red in-car glances can be therefore considered as inappropriately long in-car glances in relation to the visual demands of the given driving situation – or, in other words, visual

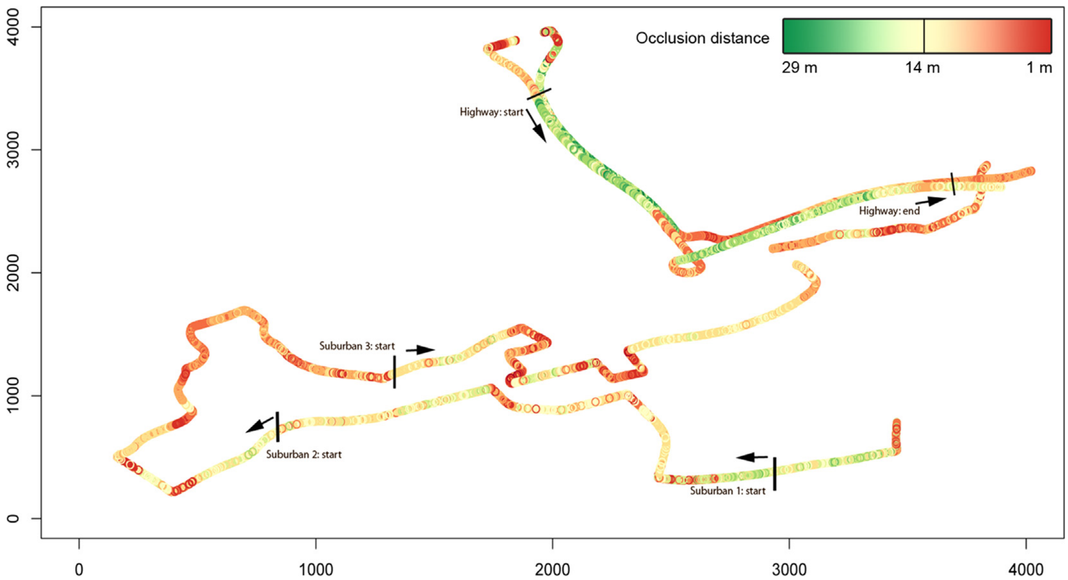

Fig. 1. Occlusion distance map (ODM) illustrates the routes used in the experiments – the color of a route point indicates the level of visual demand measured in occlusion distance (green = high occlusion distance, visually undemanding; red = low occlusion distance, visually demanding); the axes indicate route distance in meters – the figure is based on Kujala, Mäkelä, et al. (2016) and Kujala and Mäkelä (2015). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

distraction (Grahn & Kujala, 2020). According to, for instance, Wierwille (1993), drivers adapt their glancing behavior in regard to how demanding the driving scenario is. This observation fits well with the idea behind the method by Kujala and Mäkelä (2015): if we have information on how demanding a certain route point is, we could assume that the drivers should adapt their glancing behavior accordingly. If they do not, we may assume that they are distracted.

The verification threshold for the red in-car glances has been set to 6% (max) of all the in-car glances done during the testing of the task in the previous studies (Grahn & Kujala, 2020, 2018; Kujala, Grahn, et al., 2016; Kujala & Grahn, 2017). Effectively, if a participant's red in-car glance percentage exceeds the set 6%, the task is considered distractive, and the task fails the verification criterion (Kujala & Mäkelä, 2015). Kujala and Mäkelä (2015) argue for the threshold of 6% because it was the observed median of occlusion distances that exceeded the 85th percentile occlusion distances in the original experiment of 97 drivers. Differently than described in the NHTSA (2013) guidelines which are concerned with the fixed duration limits of glances dedicated to secondary tasks, the new method considers a task too distracting if more than 6% of in-car glances happen in visually demanding route points.

2.2. Experiments

The data were collected in two driving simulator experiments where participants ($N = 23 + N = 23$, no overlapping in participants) conducted the same tasks using regular Android smartphone applications. The data utilized in this study are a subset of the data originally collected for a study on comparing different user interfaces with a different scope, study angle, and research questions (Grahn & Kujala, 2020). Additionally, the new method required the use of the previously collected occlusion distances of the 97 drivers in the original experiment by Kujala, Mäkelä, et al. (2016) to check whether the distributions of the occlusion distances in the randomized samples are similar to those in that original experiment. The procedure is summarized in Fig. 2.

2.3. Design and participants

Although the data used in this study were collected for other experiments, we describe the data collection process for transparency. Our study setting was a within-subjects design for both experiments where the participants used one user interface for two in-car tasks. There were 24 participants (due to the technical problems during the testing, $N = 23$ in both analyses) in both driver samples and they were recruited using our university's mailing lists. The NHTSA (2013) recommendations on the driver sample were followed as closely as possible. Summarizing details about the participants are listed in Table 1.

The imbalance between genders was due to simulator sickness: some females had symptoms and were substituted with males. The participants were required to have driven at least 5,000 km a year. The participants read and signed an informed consent form describing the purpose of the study and data use. Before participating in the test, each participant was required

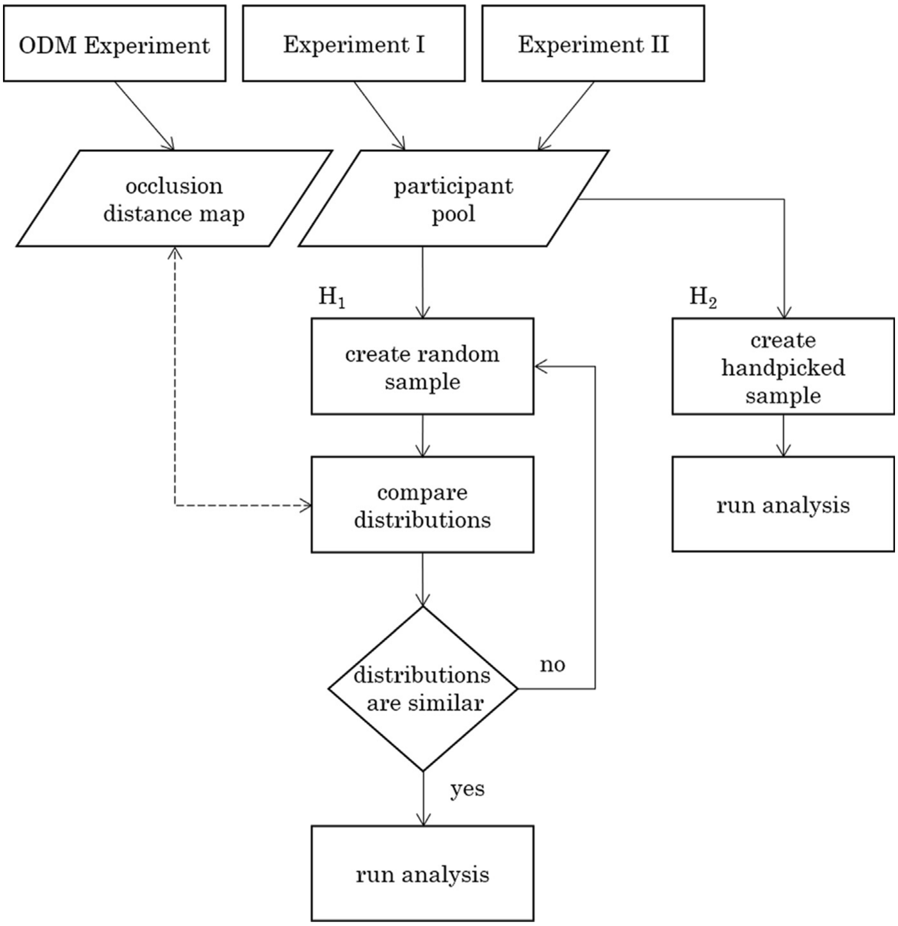

Fig. 2. The procedure – the previously conducted occlusion distance mapping (Kujala, Mäkelä et al., 2016) provides the occlusion distance map (ODM), and previously collected data for Experiments I and II (Grahn & Kujala, 2020) form the participant pool from which the samples for this study are selected – for H_1, the distributions of occlusion distances in a random sample of 23 participants was compared to the ODM, and given that the distributions were similar, the sample was selected for analysis – this was repeated until ten samples passed the distribution comparison; for H_2, the participants were handpicked based on occlusion distance.

Table 1
A summary of participants in the two experiments – these participants form the pool of participants from which the samples are selected.

	Experiment I	Experiment II
Number of participants	23	23
Number of females	7	8
Number of males	17	16
Number of participants in age group 18–24	8	7
Number of participants in age group 25–39	9	9
Number of participants in age group 40–54	4	5
Number of participants in age group 55+	3	3
Age of participants: lowest, highest	20–79 (*M* = 34.8; *SD* = 16.0)	19 – 66 (*M* = 35.3; *SD* = 13.9)
Driving experience in years: lowest, highest	2 – 55 (*M* = 16; *SD* = 15)	2 – 48 (*M* = 16.9; *SD* = 13.9)
Kilometers driven per year: lowest, highest	5,000 – 30,000 (*M* = 12,940; *SD* = 705)	5,000 – 55,000 (*M* = 14,630; *SD* = 1,185)

to have a valid driver's license, and normal or corrected-to-normal eyesight. All participants evaluated themselves as generally healthy. The experiments were instructed in Finnish and all participants understood and spoke Finnish. After the experiment, each participant was rewarded with a gift certificate (15 EUR for Experiment I and 10 EUR for Experiment II).

2.4. Apparatus

We conducted the experiments in the driving simulator laboratory of the University of Jyväskylä. In both experiments, we used a medium-fidelity driving simulator with the CKAS Mechatronics 2-DOF motion platform with Eepsoft's simulator software (http://eepsoft.fi), which saved the driving log data at 10 Hz (see Fig. 3).

The driving simulator had automatic transmission, longitudinally adjustable seat and Logitech's G27 force-feedback steering wheel and pedals. Three 40″ LED screens (95.6 cm × 57.4 cm, 1440 × 900 pixels per screen) were used to display the driving scene, a HUD RPM gauge, a HUD speedometer, a rear-view mirror (in the middle screen), and side mirrors (in side screens). For the occlusion trial, the steering wheel was equipped with two levers that displayed the driving scene for 500 ms when pulled. Otherwise, the screens were blank.

In both experiments, we used Ergoneers' Dikablis 50 Hz head-mounted eye-tracking system to record eye movements. Eye-tracking data were synchronized with driving simulator data (coordinates, speed) using a custom-built logging software and a local area network connection. Samsung Galaxy A3 smartphone (4.5″, Android 6.0.1) was utilized to run two regular Android smartphone applications: email and Spotify (see Figs. 4 and 5). The smartphone was placed in a holder next to the steering wheel (see Fig. 3) and used in portrait mode in Experiment I (see Fig. 4) and in landscape mode in Experiment II (see Fig. 5). The change of the orientation was due to the research question of the original study the data was gathered for. RStudio (version 1.0.136) and IBM SPSS Statistics 24 and 26 were used for statistical analyses.

2.5. Procedure

Before the experiment proper, the participants familiarized themselves with the driving simulator by driving in the urban environment with traffic. The participants practiced in this environment for an average time of 5.8 (Experiment I) and 4.8 min (Experiment II). Next, the participants familiarized themselves for the occlusion trial by having their vision occasionally and self-paced occluded while driving in the same urban environment. The participants practiced in this environment for an average time of 4.3 (Experiment I) and 4.0 min (Experiment II).

In both experiments, the familiarization was followed by the occlusion trial for the validation of the driver sample. As described by Senders et al. (1967), a participant's screens were blank unless the participant pulled a lever on the steering wheel, revealing the driving scene for 500 ms. Before the trial, the participants were instructed to follow driving regulations while trying to drive without visual information for as long as possible at a time. To facilitate participant's focus, a movie ticket was promised for six participants with the longest median distance accurately driven without visual information.

Fig. 3. Driving simulator and experimental setup (Experiment II).

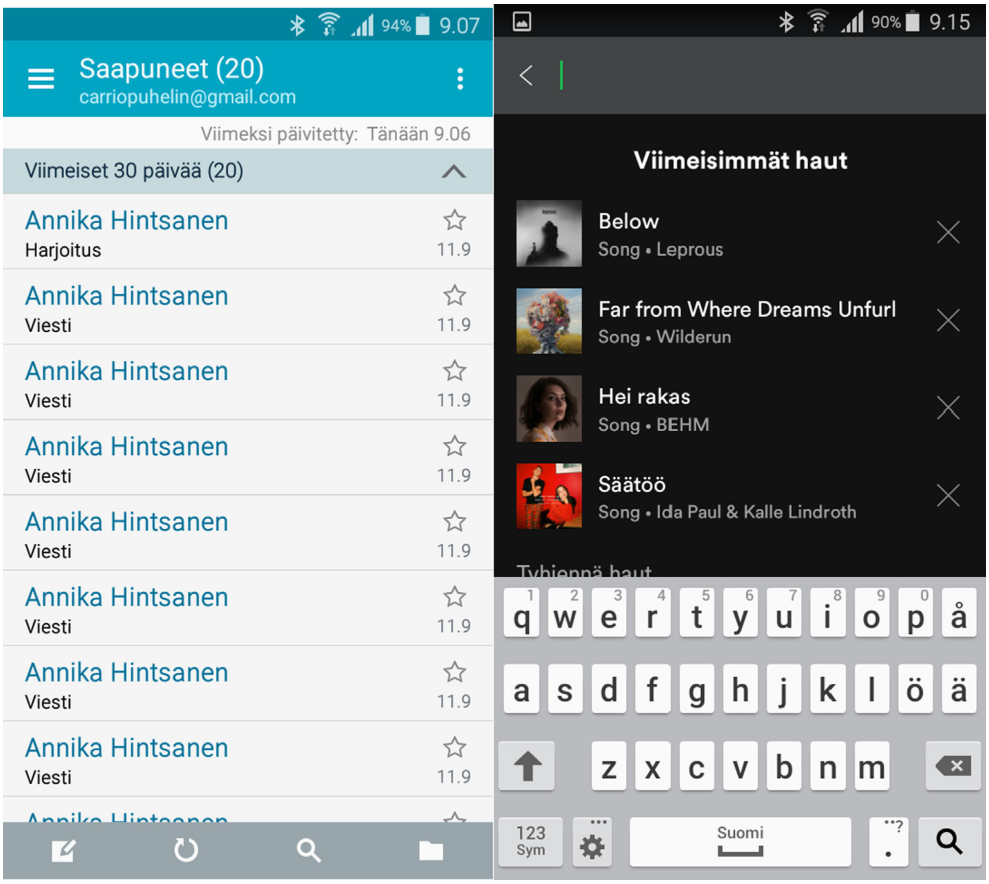

Fig. 4. Email reading and song searching in Experiment I.

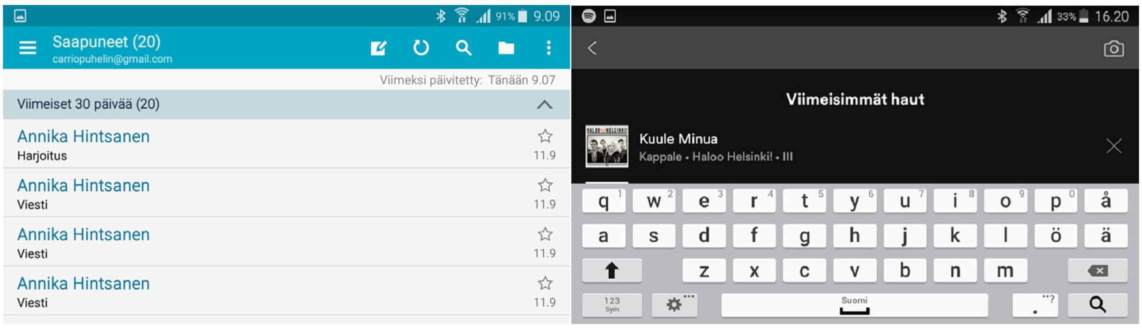

Fig. 5. Email reading and song searching in Experiment II (Grahn and Kujala, 2020).

Contrary to the familiarization before the trial, a highway environment was utilized in the trial, with speed limits of 60, 80, and 120 km/h. The participants were informed verbally when the speed limit changed. However, the speed could be adjusted if needed.

After the occlusion trial, distraction testing followed. This test utilized the head-mounted eye-tracker described earlier, and was driven in a suburban environment with a speed limit of 50 km/h. Again, speed could be adjusted if needed. In this test, the participant was required to perform tasks with the Android mobile device. Before each task, we guided the participant through the task similar to the following actual test.

In the email reading task, participants read emails and tried to find answers to four different questions we asked while the participant was driving. To be able to conduct the task, participants had to select and tap an email, read the email once it opened, and after reading, tap the back button to get back to the main email view. Participants needed to repeat this 20 times in order to read all the emails. The emails were short texts containing 104–179 characters. The task was the same in both experiments, only the orientation of the phone changed.

In the song searching task, participants used a touch screen with a *qwerty* keyboard to search a song verbally communicated by us from Spotify, start to play it, stop it, and check the artist information or check which album the song was a part of. This was repeated four times with different songs. The task was the same in both experiments, only the orientation of the phone changed. In order to prevent the learning effect, the order of the three suburban routes (Fig. 1) and tasks were counterbalanced.

2.6. Analysis

In-car glance lengths, as well as vehicle and pupil coordinates complemented by timestamps were recorded by the driving simulator and eye-tracker, and synchronized during the experiments with custom-built logging software. After the experiments, we used Noldus Observer XT software to manually check the synchronization, and corrected inaccuracies when needed. We used the SAE-J2396 (Society of Automotive Engineers, 2000) definition in scoring in-car glance lengths. As in Kujala, Grahn, et al. (2016), in-car glances that exceeded the 85th percentile of the original sample's occlusion distances (i.e., 97 drivers in Kujala, Mäkelä, et al., 2016) were labeled as red in-car glances.

3. Results

3.1. Occlusion distances: Driver sample validation

To validate the driver sample, the distributions of the occlusion distances from Experiments I and II were compared with Levene's test to the original occlusion distance distribution of 97 drivers (Kujala, Mäkelä, et al., 2016). In the original occlusion distance sample, the distances varied from 3.21 m to 41.88 m, median being 13.67. In Experiment I, the variation of occlusion distances was 6.36 to 35.82 m, median being 17.37. According to Levene's test, the variance of the occlusion distance distribution does not significantly differ from the original distribution ($F(1, 116) = 0.645$, $p = .424$). In Experiment II, the variation of occlusion distance was 4.77 to 36.00 m, median being 16.53 and the distribution does not significantly differ from the original distribution ($F(1, 117) = 0.032$, $p = .859$). In addition, as in Kujala and Grahn (2017), no association (tested with Spearman's Rank-Order Correlation) between median occlusion distances and median glance distances ($N = 46$) was found: $\rho = 0.133$, $p = .346$.

3.2. Number of in-car glances per task type

According to Kujala, Grahn, et al. (2016), the number of in-car glances should exceed 20 in order the analysis being reliable. Hence, the number of in-car glances for both task types in both experiments was sufficient for meaningful and reliable analysis (Table 2).

3.3. Red in-car glance percentages per task type

Due to non-gaussian distribution of the red in-car glances, medians were used instead of means in statistical testing. Median red in-car glances per experiment and task type are reported in Table 3.

The verification criterion for passing the distraction potential testing by Kujala and Mäkelä (2015) was set in a previous study (Kujala, Grahn, et al., 2016) to 6%, which is the maximum percentage of the red in-car glances a task can have in order to pass the distraction potential testing. If the percentage of red in-car glances exceeds 6%, it is further tested whether the difference is statistically significant. This criterion ("red in-car glance percentage equals 6") was tested with the one-sample Wilcoxon signed rank test. In Experiment I, both tasks, email reading ($Z = 3.881$, $p < .001$) and song searching ($Z = 3.716$, $p < .001$) had a red in-car glance percentage of greater than six, and differed significantly from the threshold and therefore failed the set criterion for the red in-car glances. In Experiment II, again, both tasks failed the criterion (email reading: $Z = 3.742$, $p < .001$; song searching $Z = 1.977$, $p = .048$).

Since the orientation of the smartphone used in experiments was different, therefore also the difference in the red in-car glance percentages between experiments was tested with the Mann-Whitney U test. This was done in order to test whether the orientation of the smartphone influenced the red in-car glance percentages. There was no difference between experiments in the red in-car glance percentages in either task: email reading $U = 223.000$, $p = .362$; song searching

Table 2
Mean number of in-car glances (standard deviation in parentheses).

	Email reading	Song searching
Experiment I	$M = 86.83$ (31.44)	$M = 63.78$ (23.68)
Experiment II	$M = 85.87$ (18.43)	$M = 52.78$ (14.67)

Table 3
Red in-car glance percentages (median, interquartile range in parentheses).

	Email reading	Song searching
Experiment I	*Mdn* = 19.00 (19.00)	*Mdn* = 16.00 (15.00)
Experiment II	*Mdn* = 13.93 (12.79)	*Mdn* = 8.16 (13.44)

U = 179.000, p = .060. This result supports the observations reported in Lasch and Kujala (2012), who report that there is no effect of the screen orientation on distraction.

3.4. Mixing participants: Random samples

In order to test if the driver sample affects the results of the distraction potential testing, that is, if the tasks pass the set verification criterion with a different sample of drivers, we created ten different driver samples from the total of the 23 + 23 driver samples in Grahn and Kujala (2020). We used the "random sample of cases" function of SPSS to create different driver samples. Each sample contained 23 participants. The occlusion distance distribution of each driver sample was tested with Levene's test to verify that the sample does not differ significantly from the original sample of Kujala, Mäkelä, et al. (2016). Since one of the distributions statistically differed from the original occlusion distance distribution, it was omitted from the testing and replaced with a new one. Then, the one-sample Wilcoxon signed rank test was used to test if the red in-car glance percentages pass the set criterion (max. 6%) and differ significantly. With each driver sample, both tasks failed the set verification criterion: the red in-car glance percentage was over 6, and the percentage differed significantly from the threshold. The results of the distraction potential testing and descriptive statistics of the samples are reported in Table 4.

3.5. Mixing participants: Occlusion distance

To test the effect of occlusion distances on the results of the distraction potential testing, we handpicked a sample that consisted of drivers with median occlusion distance less than or equal to 16 m (N = 18). With the driver sample of median occlusion distance less than or equal to 16 m, the song searching task passed the set verification criterion: median red in-car glance percentage was 8.08% but it did not significantly differ from the threshold of 6% (p = .117). The email reading task did not pass the set verification criterion with this driver sample either. The results of the distraction potential testing and descriptive statistics of the sample are reported in Table 5.

In summary, both hypotheses were supported. For H_1, none of the results performed using the ten random samples differed from the results reported in Grahn and Kujala (2020). For H_2, handpicking a sample of drivers with low occlusion distance changed the result of distraction potential testing.

4. Discussion

Previous studies have indicated that drivers have individual glancing behaviors while conducting secondary in-car tasks (e.g., Broström et al., 2013; Donmez et al., 2010; Kujala et al., 2014; Kujala & Grahn, 2017), and these individual differences may affect the results of the distraction potential testing (Broström et al., 2016; J. Y. Lee & Lee, 2017; Ljung Aust et al., 2015). Individual differences in glancing behaviors are not considered, for example, in commonly known NHTSA's (2013) Driver Distraction Guidelines for In-Vehicle Electronic Devices. In this study, we tested if a new, occlusion distance-based distraction potential testing method (Kujala & Mäkelä, 2015) better accounts for individual glancing behaviors than the NHTSA method. To that end, we mixed participants from two driving simulator experiments and tested if the results of the distraction potential testing changed when compared to previously reported results with the same dataset of Grahn and Kujala (2020).

Table 4
Results of the ten random participant samples; OD refers to occlusion distance.

Sample #	Levene's test (H_0: OD distribution differs from the original N = 97 OD distribution)	Median red in-car glance percentage and the result of the distraction potential testing (email reading)	Median red in-car glance percentage and the result of the distraction potential testing (song searching)	Occlusion distance (median)	Occlusion distance range (median)	Age (mean, *SD* in parentheses)	Age range
1	F = 0.025, p = .874	19.00%, Z = 3.894, p < .001, fail	8.16%, Z = 2.738, p = .006, fail	17.31	6.35–35.82	33.7 (14.46)	19–76
2	F = 0.007, p = .934	16.00%, Z = 3.893, p < .001, fail	10.42%, Z = 2.768, p = .006, fail	19.66	6.39–35.99	32.7 (12.78)	20–65
3	F = 0.822, p = .366	13.48%, Z = 3.362, p = .001, fail	11.00%, Z = 3.228, p = .001, fail	16.71	6.35–33.96	34.6 (12.09)	21–76
4	F = 0.480, p = .490	17.57%, Z = 3.590, p < .001, fail	14.00%, Z = 2.921, p = .003, fail	16.64	6.35–35.99	30.2 (12.51)	21–76
5	F = 2.884, p = .092	16.39%, Z = 3.909, p < .001, fail	13.95%, Z = 2.952, p = .003, fail	17.31	6.39–35.82	33.1 (12.30)	19–65
6	F = 0.022, p = .883	16.00%, Z = 3.772, p < .001, fail	16.00%, Z = 3.772, p < .001, fail	17.31	6.35–35.82	32.5 (12.49)	21–76
7	F = 0.002, p = .962	17.57%, Z = 3.529, p < .001, fail	14.00%, Z = 3.164, p = .002, fail	17.31	6.35–35.82	34.5 (14.34)	19–76
8	F = 0.575, p = .450	13.92%, Z = 3.773, p < .001, fail	11.00%, Z = 3.194, p = .001, fail	16.64	6.39–33.96	31.6 (10.99)	21–60
9	F = 0.490, p = .485	16.00%, Z = 3.772, p < .001, fail	14.00%, Z = 3.286, p = .001, fail	20.61	6.35–35.99	36.6 (13.07)	21–76
10	F = 0.403, p = .527	13.93%, Z = 3.665, p < .001, fail	8.82%, Z = 2.433, p = .015, fail	17.03	6.39–33.96	34.7 (12.73)	20–65

Table 5
Results of participant sample with occlusion distance (OD) ≤ 16 m.

Sample #	Levene's test (H_0: OD distribution differs from the original *N* = 97 OD distribution)	Median red in-car glance percentage and the result of the distraction potential testing (email reading)	Median red in-car glance percentage and the result of the distraction potential testing (song searching)	Occlusion distance (median)	Occlusion distance range	Age (mean, *SD* in parentheses)	Age range
OD ≤16	F = 10.784, p = .001	16.20%, Z = 2.919, p = .004, fail	8.08%, Z = 1.568, p = .117, pass (d = 0.545, medium effect)	13.47	6.35–16.03	34.7 (15.91)	21–76

Based on the ten different mixed driver samples, the tasks labeled distractive remained distractive, regardless of the driver sample. This gives support for H_1: *The results of the new occlusion distance-based distraction potential testing method do not change when the sample of participants is randomized.* This implies that the new distraction potential testing method introduced by Kujala and Mäkelä (2015) considers drivers' individual differences in glancing behavior – potentially describing the phenomenon more accurately than the NHTSA (2013) method. However, as discussed earlier, the new method requires the occlusion distances to be mapped prior to the experiments. This can be seen as a drawback when compared to the NHTSA method, as the new method requires an additional experiment in order the produce the occlusion distance map, before the experiments proper can be implemented. However, given that the new method arguably also captures a more realistic driving scenario, this additional work may be one way towards more accurate results.

As previously shown (Ljung Aust et al., 2015), the results of the NHTSA (2013) method can be manipulated by varying the participant sample. For instance, Broström et al. (2016) and Ljung Aust et al. (2015) managed to affect the results of the distraction potential testing following the NHTSA guidelines and by re-selecting the driver samples. In the NHTSA method, the important part of the method is the recommendations of driver sample regarding the ages of participants. The ages of participants are important since it is known that higher age implies longer durations of in-car glances (e.g., Dobres et al., 2016; J. Lee et al., 2015; Son & Park, 2012; Wikman & Summala, 2005). However, the studies of Broström et al. (2016) and Ljung Aust et al. (2015) did not report the ages of the drivers in their manipulated driver samples, and therefore the effects of age cannot be evaluated in their studies. However, in this study, we had both younger and older drivers (see *age range* in Tables 4 and 5). In addition to driver age, we had information on the drivers' median occlusion distances, and we ensured that each driver sample contained drivers with varying median occlusion distances (see *occlusion distance* in Table 4).

A rather interesting observation was the effect of occlusion distance on distraction potential testing: with a driver sample with median occlusion distance lower than or equal to 16 m, while still including both younger and older drivers (Table 5), we were able to produce a contrary test result where the task was not considered distracting. This observation supports H_2: *The results of the new occlusion distance-based distraction potential testing method change when only participants with low occlusion distance (Mdn ≤ 16 m) are selected*. In the sample, there were only drivers who were able to drive less than or equal to 16 m (median) with occluded vision in the occlusion trial, and after 16 m (median) they felt they needed visual information to maintain safe driving. In other words, participants who had low occlusion distance had less inappropriately long in-car glances (red in-car glances) during the distraction testing. These drivers could have been insecure regarding their driving skills and therefore tried to keep their in-car glances as short as possible. In addition, the structure of the tested task may have allowed the drivers to use subtask boundaries as natural break points during the task completion (Janssen et al., 2012; J. Y. Lee et al., 2015; J. Y. Lee & Lee, 2019; Salvucci & Kujala, 2016) in order to avoid inappropriately long in-car glances. However, the glancing behavior of drivers with low occlusion distance resulted in passing the distraction testing – even when it failed with other randomly mixed samples which had similar occlusion distance distribution as in the original occlusion distance study of *N* = 97 (Kujala, Mäkelä, et al., 2016). Finally, this observation indicates that it is possible, even likely, that a distractive in-car task passes distraction potential testing due to the neglection of drivers' individual differences in glancing behaviors.

The observation that a task passes the distraction testing with an all-aged driver sample with low occlusion distance is rather interesting. Previous studies have concluded that in-car glance durations increase with age. In our experiment, the age range of the driver sample with median occlusion distance lower than or equal to 16 m was 21–76 years, mean age being 35.5 years. That is to say, the sample also included younger drivers. This observation suggests that age is not the only factor affecting glance durations, and thus including older drivers in the driver samples (as in the NHTSA method) is by itself not a sufficient approach in taking drivers' individual glancing behaviors into account.

Overall, these findings suggest that a new distraction testing method (Kujala & Mäkelä, 2015) accounts for drivers' individual glancing behaviors and therefore may produce more robust distraction testing results regarding driver selection. Based on our results, a participant's occlusion distance may have an association with inappropriately long in-car glances during the distraction testing, and when conducting distraction potential testing, driver samples should be validated with their occlusion distances to ensure that the individual glancing behaviors have been considered.

It should be noted that the orientation of the smartphone was different between the experiments. In Experiment I, the smartphone was in portrait mode, and in Experiment II in landscape mode. This was due to the research questions and research angle this data were initially collected. However, the original results between orientations did not differ significantly (as reported in Section 3.3), and previously, it has been discovered that the orientation of the used device does not affect distraction potential testing (Lasch & Kujala, 2012). It should also be noted that the red in-car glance percentages were relatively high in the original experiments of Grahn and Kujala (2020), and this could have affected the results of the mixed driver samples in this paper despite the individual glancing behaviors. However, despite the high red in-car glance percentages, we were able to produce a result where the task passed the distraction potential testing.

Furthermore, this driver sample mix could be further replicated with tasks that have lower red in-car glance percentages. Another consideration for future research is a comparison between the results of the new distraction testing method (Kujala & Mäkelä, 2015) and the NHTSA (2013) method. One possibility for such a research setting is to study a phenomenon first using the NHTSA scenario, and then using the scenario described in the new method (or vice versa), and finally comparing the results. Finally, although our results indicate that occlusion distance may be a more accurate validation criterion for driver sampling than merely age, a number of factors still remain unstudied in this context, for example, the association between occlusion distances and Attention-Related Driving Errors Scale (Cheyne et al., 2006), as well as Hazard Prediction Test (Crundall, 2016).

5. Conclusion

In this study, we set out to investigate if by manipulating driver samples we can affect distraction testing results when using a new, previously reported distraction potential testing method that validates the driver sample based on drivers' occlusion distances. This validation ensures that the driver sample contains drivers with different glancing behaviors measured with occlusion distance. Our results indicate that these results obtained with this new method are not affected by manipulating driver samples when the sample includes all kinds of drivers – from those who are able to drive longer occlusion distances to those who are able to drive shorter occlusion distances. This indicates that the method tested in this study might account for individual driver differences more accurately than, for example, tests that utilize the NHTSA guidelines, which are shown to be susceptible to participant sample manipulation. Effectively, this could mean that just leaning on the assumption that including older drivers in the sample ensures that the sample contains drivers with different glancing behaviors, and thus considers individual differences, may not be accurate. Hence, without accounting for individual glancing behaviors validated with occlusion distances, there is a potential for false passing of distraction tests. These empirical findings may be utilized in refining the existing guidelines, thus providing increased scientific rigor in distraction potential testing.

CRediT authorship contribution statement

Hilkka Grahn: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing - original draft. **Toni Taipalus:** Writing - review & editing.

Acknowledgments

We would like to thank the Associate Editor and the anonymous Reviewers for their valuable comments and suggestions on how to improve the paper, and Tuomo Kujala for his valuable comments.

References

Bayer, J. B., & Campbell, S. W. (2012). Texting while driving on automatic: Considering the frequency-independent side of habit. *Computers in Human Behavior, 28*(6), 2083–2090. https://doi.org/10.1016/J.CHB.2012.06.012.

Broström, R., Aust, M. L., Wahlberg, L., & Källgren, L. (2013). What drives off-road glance durations during multitasking – capacity, practice or strategy?. *Proceedings on the 3rd International Conference on Driver Distraction and Inattention*.

Broström, R., Bengtsson, P., & Aust, M. L. (2016). Individual glance strategies and their effect on the NHTSA visual manual distraction test. *Transportation Research Part F: Traffic Psychology and Behaviour, 36*, 83–91. https://doi.org/10.1016/J.TRF.2015.10.017.

Buchhop, K., Edel, L., Kenaan, S., Raab, U., Böhm, P., & Isemann, D. (2017). In-vehicle touchscreen interaction: Can a head-down display give a heads-up on obstacles on the road? In *Proceedings of the 9th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 21–30). https://doi.org/10.1145/3122986.3123001.

Caird, J. K., Johnston, K. A., Willness, C. R., Asbridge, M., & Steel, P. (2014). A meta-analysis of the effects of texting on driving. *Accident Analysis and Prevention, 71*, 311–318. https://doi.org/10.1016/j.aap.2014.06.005.

Cheyne, J. A., Carriere, J. S. A., & Smilek, D. (2006). Absent-mindedness: Lapses of conscious awareness and everyday cognitive failures. *Consciousness and Cognition, 15*(3), 578–592. https://doi.org/10.1016/j.concog.2005.11.009.

Choudhary, P., & Velaga, N. R. (2017). Modelling driver distraction effects due to mobile phone use on reaction time. *Transportation Research Part C: Emerging Technologies, 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. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 1693–1697*. https://doi.org/10.1177/1071181312561339.

Crundall, D. (2016). Hazard prediction discriminates between novice and experienced drivers. *Accident Analysis and Prevention, 86*, 47–58. https://doi.org/10.1016/j.aap.2015.10.006.

Dobres, J., Reimer, B., Mehler, B., Foley, J., Ebe, K., Seppelt, B., & Angell, L. (2016). The influence of driver's age on glance allocations during single-task driving and voice vs. visual-manual radio tuning. SAE Technical Papers. https://doi.org/10.4271/2016-01-1445

Domeyer, J. E., Diptiman, T., & Hamada, H. (2014). Using occlusion to measure the effects of the NHTSA participant criteria on driver distraction testing. In *Proceedings of the Human Factors and Ergonomics Society* (pp. 2205–2209). https://doi.org/10.1177/1541931214581463.

Donmez, B., Boyle, L. N., & Lee, J. D. (2010). Differences in off-road glances: Effects on young drivers' performance. *Journal of Transportation Engineering, 136* (5), 403–409. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000068.

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? *Computers in Human Behavior, 76*, 174–183. https://doi.org/10.1016/J.CHB.2017.07.023.

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*. https://doi.org/10.1016/j.ijhcs.2020.102467.

Grahn, H., & Kujala, T. (2018). Visual distraction effects between in-vehicle tasks with a smartphone and a motorcycle helmet-mounted head-up display. *ACM International Conference Proceeding Series, 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. *Transportation Research Record: Journal of the Transportation Research Board, 2147*(1), 66–74. https://doi.org/10.3141/2147-09.

He, J., Chaparro, A., Wu, X., Crandall, J., & Ellis, J. (2015). Mutual interferences of driving and texting performance. *Computers in Human Behavior, 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. (2015). Texting while driving using Google Glass[TM]: Promising but not distraction-free. *Accident Analysis and Prevention, 81*, 218–229. https://doi.org/10.1016/j.aap.2015.03.033.

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. *Journal of Cognitive Engineering and Decision Making, 6*(1), 5–29. https://doi.org/10.1177/1555343411432339.

Kircher, K., Kujala, T., & Ahlström, C. (2019). On the difference between necessary and unnecessary glances away from the forward roadway: An occlusion study on the motorway. *Human Factors*. https://doi.org/10.1177/0018720819866946.

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. (2016). 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. (2015). Testing environment and verification procedure for in-car tasks with dynamic self-paced driving scenarios. *Proceedings of the 4th International Conference on Driver Distraction and Inattention*.

Kujala, T., Mäkelä, J., Kotilainen, I., & Tokkonen, T. (2016). The attentional demand of automobile driving revisited: Occlusion distance as a function of task-relevant event density in realistic driving scenarios. *Human Factors, 58*(1), 163–180. https://doi.org/10.1177/0018720815595901.

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., Reimer, B., Mehler, B., Angell, L., Seppelt, B. D., & Coughlin, J. F. (2015). Analyses of glance patterns of older and younger drivers during visual-manual human-machine interface interaction. In *Transportation Research Board 94th Annual Meeting*. https://doi.org/10.1021/es501173p.

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., & Lee, J. D. (2019). Modeling microstructure of drivers' task switching behavior. *International Journal of Human-Computer Studies, 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. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 1919–1923*. https://doi.org/10.1177/1541931213601960.

Ljung Aust, M., Rydstöm, A., Broström, R., & Victor, T. (2015). Better ways to calculate pass/fail criteria for the eye glance measurement using driving simulator test. *4th Inernational Driver Distraction and Inattention Conference*.

Milgram, P. (1987). A spectacle-mounted liquid-crystal tachistoscope. *Behavior Research Methods, Instruments, & Computers.*. https://doi.org/10.3758/BF03205613.

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

Ng, A., & Brewster, S. (2017). An evaluation of touch and pressure-based scrolling and haptic feedback for in-car touchscreens. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 11–20). https://doi.org/10.1145/3122986.3122997.

Oviedo-Trespalacios, O., Haque, M. M., King, M., & Washington, S. (2016). Understanding the impacts of mobile phone distraction on driving performance: A systematic review. *Transportation Research Part C: Emerging Technologies, 72*, 360–380. https://doi.org/10.1016/J.TRC.2016.10.006.

Perlman, D., Samost, A., Domel, A. G., Mehler, B., Dobres, J., & Reimer, B. (2019). The relative impact of smartwatch and smartphone use while driving on workload, attention, and driving performance. *Applied Ergonomics, 75*, 8–16. https://doi.org/10.1016/j.apergo.2018.09.001.

Regan, M. A., Hallett, C., & Gordon, C. P. (2011). Driver distraction and driver inattention: Definition, relationship and taxonomy. *Accident Analysis and Prevention, 43*(5), 1771–1781. https://doi.org/10.1016/j.aap.2011.04.008.

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. In MIT AgeLab Technical Report No. 2013-17A. Massachusetts Institute of Technology, Cambridge, MA. https://doi.org/2013-18A

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. *Accident Analysis and Prevention, 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. 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. *Highway Research Record*.

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

Son, J., & Park, M. (2012). Comparison of younger and older drivers' glance behavior and performance in a driving simulator. *25th ARRB Conference*.

Tivesten, E., & Dozza, M. (2014). Driving context and visual-manual phone tasks influence glance behavior in naturalistic driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 26*, 258–272. https://doi.org/10.1016/j.trf.2014.08.004.

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

Tsimhoni, O., & Green, P. (2001). Visual demand of driving and the execution of display-intensive in-vehicle tasks. *Proceedings of the Human Factors and Ergonomics Society, 1586–1590*. https://doi.org/10.1177/154193120104502305.

Villalobos-zúñiga, G., Kujala, T., & Oulasvirta, A. (2016). T9 + HUD : Physical keypad and HUD can improve driving performance while typing and driving. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 177–184). https://doi.org/10.1145/3003715.3005453.

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.

Wikman, A. S., & Summala, H. (2005). Aging and time-sharing in highway driving. *Optometry and Vision Science, 82*(8), 716–723. https://doi.org/10.1097/01.opx.0000174721.33377.b8.