Author's Accepted Manuscript

CONTEXT-SENSITIVE DISTRACTION WARNINGS – EFFECTS ON DRIVERS' VISUAL BEHAVIOR AND ACCEPTANCE

Tuomo Kujala, Hannu Karvonen, Jakke Mäkelä

www.elsevier.com/locate/ijhcs

PII: S1071-5819(16)00035-5
DOI: http://dx.doi.org/10.1016/j.ijhcs.2016.03.003
Reference: YIJHC2017

To appear in: *Journal of Human Computer Studies*

Received date: 1 June 2015
Revised date: 1 March 2016
Accepted date: 6 March 2016

Cite this article as: Tuomo Kujala, Hannu Karvonen and Jakke Mäkelä CONTEXT-SENSITIVE DISTRACTION WARNINGS – EFFECTS ON DRIVERS' VISUAL BEHAVIOR AND ACCEPTANCE, *Journal of Huma Computer Studies,* http://dx.doi.org/10.1016/j.ijhcs.2016.03.003

CONTEXT-SENSITIVE DISTRACTION WARNINGS – EFFECTS ON DRIVERS' VISUAL BEHAVIOR AND ACCEPTANCE

Tuomo Kujala*, Hannu Karvonen¶, Jakke Mäkelä*

*University of Jyväskylä

Department of Computer Science and Information Systems

P.O. Box 35

FI-40014 University of Jyväskylä

Finland

Corresponding author: Tuomo Kujala

tuomo.kujala@jyu.fi

jakke.makela@gmail.com

tel. +358 40 024 7392, fax +358 14 260 4400

¶VTT Technical Research Centre of Finland Ltd

P.O. Box 1000

FI-02044 VTT

Finland

hannu.karvonen@vtt.fi

Keywords: driver distraction; smart phone; warning system; situation awareness; acceptance; trust

Abstract

In this study, we investigated the effects of context-sensitive distraction warnings on drivers' in-car glance behaviors and acceptance. The studied prototype warning application functions on a smart phone. The novelty of the application is its proactive and context-sensitive approach to the adjustment of warning thresholds according to the estimated visual demands of the driving situation ahead. In our study, novice and experienced drivers conducted in-car tasks with a smart phone on a test track with and without the warnings. The application gave a warning if the driver's gaze was recognized to remain on the smart phone over a situation-specific threshold time, or if the driver was approaching a high-demand part of the track (an intersection or a tight curve). Glance metrics indicated a significant increasing effect of the warnings on glance time on road while multitasking. The effect varied between 5 to 30% increase depending on the in-car task. A text message reading task was the most visually demanding activity and indicated the greatest effect of the warnings on glance time on road. Driving experience did not have an effect on the efficiency of the warnings. The proposed gaze tracking with current smart phone technology proved to be highly unreliable in varying lighting conditions. However, the findings suggest that location-based proactive distraction warnings of high-demanding driving situations ahead could help all drivers in overcoming the inability to evaluate situational demands while interacting with complex in-car tasks and to place more attention on the road. Furthermore, survey results indicate that it is possible to achieve high levels of trust, perceived usefulness, and acceptance with these kinds of context-sensitive distraction warnings for drivers.

1 INTRODUCTION

Modern smart phones offer car drivers a lot of useful services on the road such as navigation, entertainment, communication, and information on nearby points-of-interest. However, a concern has been addressed lately on the increasing smart phone usage while driving and the related inattention towards the traffic environment (Fitch, Soccolich, Guo, F., et al., 2013; Klauer, Dingus, Neale, Sudweeks, and Ramsey, 2006).

From earlier research it is known that driver inattention is a major cause of safety-critical incidents in traffic. In a naturalistic driving study with one hundred car drivers (Klauer et al., 2006), it was concluded that almost 80 percent of all crashes and 65 percent of all near-crash situations involved visual inattention, i.e., the driver's eyes weren't on the road the moment before or at the moment of the incident.

As a cause of visual inattention by secondary activities in these safety-critical events, the use of a mobile device (mainly mobile phone) was by far the leading factor by at least 30% of the cases (Klauer et al., 2006). Another naturalistic driving study on the topic by Fitch et al. (2013) indicated that drivers engaging in visually complex tasks with their smart phones have a three-time higher safety-critical incident risk compared to drivers who pay attention to the road ahead.

Unfortunately, the most obvious solution to the problem, legislative measures, does not seem to work. For instance, in Finland a recent poll by the Finnish Road Safety Council revealed that over 30 percent of drivers admit texting while driving, despite of the fees on hand-held device usage and distracting in-car activities while driving (Jääskeläinen and Pöysti, 2014). This means that there is an urgent need for other, more effective means to mitigate the negative effects of driver distraction by mobile devices. Other possible approaches are, for instance, driver education and technological counter-measures. In order to provide efficient counter-measures, the priority should be on means that are widely accepted by the drivers (Donmez, Boyle, and Lee, 2007). In this paper, we study the efficiency and acceptability of context-sensitive distraction warnings that could serve this purpose.

2 DISTRACTION ALGORITHMS AND DRIVER ACCEPTANCE

Due to the increasing significance of driver distraction to traffic safety, a number of distraction detection algorithms and distraction warning systems are currently under development by car manufacturers (NHTSA, 2013a; Lee, Moeckli, Brown et al., 2013). These warning systems operate on the basis of distraction detection algorithms, i.e., algorithms that are meant to detect when the driver is distracted. However, there are basic conceptual difficulties in defining and operationalizing accurately what is distracted (inattentive) driving (Regan, Hallett, and Gordon, 2011). This places great challenges for the sensitivity and reliability of the algorithms in detecting distracted driving, and consequently, to drivers' acceptance of the distraction warnings.

Liang, Lee, and Yekhshatyan (2012) studied 24 different possible algorithms that could be used for detecting distraction and evaluated their ability to predict crash risk based on behavioral data collected in the 100 car study by Klauer et al. (2006). They concluded that the most sensitive indicator for crash risk seemed to be algorithms that measure instantaneous changes in off-road glance duration, that is, individual glance durations seem to matter. 1.5th power of glance duration, glance history, or glance location, did not significantly improve the sensitivity.

Even if the algorithms are highly valuable for indicating the general statistical link between off-road glance durations and crash risk, environmental and external situational factors (e.g., driving speed, road curvature and road type) were missing in all of the evaluated 24 algorithms (Liang et al., 2012). That is, one can argue that the severity of an off-road glance duration should be in a relationship with the visual demands of the driving situation, as suggested by the naturalistic driving study of Tivesten and Dozza (2014) as well as the 100 car study report by Klauer et al. (2006). Taking into account the situational visual demands of the driving task could further improve the sensitivity of the single glance algorithms.

The existing and proposed distraction warning systems and detection algorithms do not utilize context and driver data to the extent that could be possible with modern technology. Instead, the algorithms focus only on off-road glance durations and the direction of gaze (NHTSA, 2013a). Context-sensitivity of distraction warning systems could decrease substantially the high levels of false alarms experienced with the current systems (NHTSA, 2013a). In addition, context-sensitivity could improve the visibility of the system behavior by providing the driver a possibility to better associate the criticality of the warnings to the observable demands in the driving environment (e.g., an intersection ahead). All these factors should increase driver acceptance of these systems and make the systems more reliable. In addition, positive learning effects could be expected if the driver learns to associate the warnings to certain driving environments or situations observed ahead.

Like other available technical solutions to mitigate the negative effects of driver distraction, such as braking and lane-keeping assistants, most of the distraction warning systems today are reactive, that is, the systems react to observed distraction or its negative effects by counter-measures (e.g., Wege and Victor, 2014; You, Montes-de-Oca, Bao1 et al., 2012). This means often already a degraded driving performance.

Other, somewhat context-sensitive counter-measures act as workload managers, limiting the access of drivers to certain in-car services when the situational demands are considered to reach a certain level of high demand (Green, 2004). These kinds of forced solutions are rarely well accepted by the drivers. In addition, the high workload conditions are often recognized based on the high levels of activity by the driver (e.g., steering frequency, Green, 2004; Broström, Engström, Agnvall, and Markkula, 2006), whereas lack of sufficient attention on the driving task manifests often as low levels of activity compared to what the situational driving task demands would require (Regan et al., 2011).

The ideas about drivers themselves acting as dynamic workload managers and driver assistant systems for this purpose are relatively new (Donmez, Boyle, and Lee, 2008). A basic requirement for this kind of tactical behavior is that the driver is capable to evaluate the dynamic demands of each driving situation ahead. In-car tasks undermine this ability because it has been shown that drivers can have a low level of awareness of their own performance as well as the elements in the road environment while multitasking (e.g., Schömig and Metz, 2013; Young, Salmon, and Cornelissen, 2013; Horrey, Lesch, and Garabet, 2009). For instance, Young and Salmon (2012) have suggested that high levels of cognitive workload due to in-car task demands can have a negative effect on driver's situation awareness of the road environment, which could at least partially explain this inability. The study by Lee, Lee, and Boyle (2007) indicated that brief glances off road together with cognitive load are additive in their effects on drivers to miss safety-critical events in the driving environment.

In addition, even if the drivers would be aware of the situational driving demands, the most popular survival strategy in multitasking while driving seems to be "ASAP"; the in-car task is completed as soon as possible without considering the situational driving demands. Horrey and Lesch (2009) showed that although drivers seemed to be aware of the demands of the driving situation in their experiment, the drivers did not tend to postpone the presented secondary tasks even if they were given the chance. Based on the findings, the authors suggested that training drivers on tactical decisions and planning of timing in multitasking is worth considering. The effects of this type of training of tactical and strategic skills has been tested by Horrey et al. (2009), giving promising results. Another possibility is to provide real-time feedback for the drivers (Donmez, Boyle, and Lee, 2007), or both real-time and retrospective feedback (Roberts, Ghazizadeh, and Lee, 2012) on distracted behaviors. The study by Donmez, Boyle, and Lee (2010) indicated the positive effects of combined real-time and retrospective feedback on distracted driving behaviors among young high-risk drivers, in particular. Roberts, Ghazizadeh, and Lee (2012) suggested that systems providing immediate feedback on distracted behaviors are experienced in general as less pleasant and less easy to use than retrospective feedback systems. However, the specific implementation of the warnings can be argued to have a significant effect on the acceptability of the real-time warnings.

Instead of mere feedback, one possibility is to give the drivers proactive suggestions to postpone in-car tasks if the driving situation ahead is recognized as high demanding. A proactive and context-sensitive distraction warning system that would adjust warning thresholds according to the expected visual demands of the driving situation ahead and indicate these in real-time for the driver could in principle answer the issues raised by earlier research. In this paper, we study one possible implementation of such a prototype system called VisGuard ("Vision Guard", Kujala, 2013).

3 VISGUARD: PROTOTYPE FEATURES

In order to study the effects of context-sensitive distraction warnings on the drivers' visual behaviors and driver acceptance, we developed an Android-based mobile application called VisGuard (Kujala, 2013). The VisGuard prototype application displays the warnings on the smart phone that the driver is using while driving (see Figure 1). The application is intended to work proactively; warning the driver of the usage of the phone already before the driver enters a visually highly demanding situation.

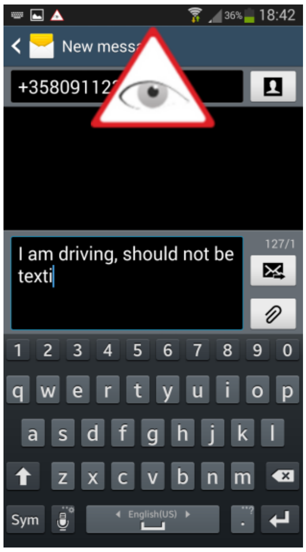

Figure 1. The warning icon on a smart phone screen.

The VisGuard prototype works as a background process in an Android smart phone. It constantly monitors the GPS, and uses a special-purpose map to estimate the visual demand level of driving at a given location on the road. The prototype was built using open source software and open data for the main tasks: Open Street Map for the visual demand map and Open CV software for gaze recognition. Near Field Communication (NFC) with a Radio-Frequency Identification (RFID) tag can be used to automatically activate the software once the driver places the device into a dashboard holder. For hand-held use, Android's activity recognition API is used to detect when the device is being used in a car. The software has been verified on Samsung Galaxy S3 and Note 2, and may work also on other high-end Android smart phones running Android 4.2 or later.

The visual demand algorithm behind the calculation of the adaptive situational warning time is based on a driving simulator study with visual occlusion and 97 drivers by Kujala, Mäkelä, Kotilainen, and Tokkonen (2016). The data allowed us to identify visually high-demanding driving scenarios in which visual distraction would be potentially particularly dangerous. It also allowed us to adjust the warning thresholds for in-car glance durations by situational and driver-specific factors.

The visual demand level is determined based on the experience level of the driver, the proximity of intersections, junctions and pedestrian crossings ahead, the winding of the road ahead, as well as the speed of the car. In addition, other forms of data, such as headway distance, traffic, weather, visibility, and device input data, may be be utilized in future implementations of the application.

The warning threshold is expressed here as a Warning Time (WT): a situation-specific threshold for a single in-car glance duration above which off-road glances are considered as risky. This varies from 0 to 2.0 seconds and corresponds to the time it takes to travel with the driver-selected speed the distance the 85th percentile of drivers preferred to travel occluded (i.e., blinded) when fully concentrating on driving in the study of Kujala et al. (2016). The use of the 85th percentile is based on a common standard in traffic engineering (TRB, 2003), assuming that 15 percent of drivers represent risky or unacceptable behaviors in traffic. The curvature of the road ahead is taken into account in our algorithm as OD = 85th percentile OD - 9.1*W, where OD is the occlusion distance in meters and W Winding as meter/meter, a measure of how much the driver needs to turn the steering wheel while driving along the road, based on Kujala et al. (2016). The areas with 85th percentile ODs of less than 5 meters (e.g., in intersections or tight curves) are considered to be visually high-demanding, and the WT is set to zero. When VisGuard determines that the threshold is exceeded or that the WT=0 (a highly demanding area ahead), it shows a warning icon (Figure 1) on the screen, which tells the driver to focus on the driving scene. For the areas of WT=0, the borders of the area are calculated based on vehicle speed and the corresponding stopping distance to the middle point of the area.

The application monitors the face of the driver in real-time, and aims to identify when the driver is looking at the phone. The gaze recognition is technically based on the phone's front camera, Open CV face recognition software, and the angle of the driver's face from the camera. For optimal gaze recognition for our experiment, the smart phone was placed on a dashboard holder so that the driver's face was not recognized when the driver was looking ahead on the road. It is known that people prefer to move their heads, instead of eyes only, if the target is more than 30 degrees away from the line of sight (Flannagan and Sivak, 1993). Therefore, a placement of the holder with at least a 30 degree angle to the driver was preferred for the testing purposes (as in Figure 2). The negative side effect of this placement is that it will recognize gaze to the device whenever the driver is looking at the right-hand field of view, but it was assumed that the driver ignores what is on the display when looking elsewhere in the driving environment. No sound or vibration was used by default, because of the possibility of false alarms and the consequent possibility to capture the driver's attention unnecessarily in these situations.

Figure 2. The phone holder installed on the dashboard air vent.

Grey circle symbol on the display indicated for the driver that gaze at the device is not recognized and the fulfillment level of the circle indicated the WT for the current situation (see Figure 3). This was intended to provide the driver a chance to check the WT with a very brief glance (less than 500 ms) before the gaze tracking catches the gaze in order to support the assessment of the expected visual demands of the situation ahead. Empty circle indicated a 2.0 s warning threshold and a half circle a 1.0 s threshold. VisGuard started a countdown as soon as a glance to the phone is identified. This was indicated by changing the color of the circle symbol to orange and showing a counter inside the circle symbol (the circle filling up, see Figure 3). In controlled laboratory settings the mean delay in gaze recognition for Samsung Galaxy S3 was near 500 ms (RMSE=521 ms, N=49 glances) and the delay was taken into account in the countdown. The orange symbol itself was intended to serve as a reminder of a potentially dangerous activity and to remind the driver to look back at the road before the orange circle was full. When a risky glance duration was observed by the application (i.e., the circle had filled up), the warning icon was displayed (see Figure 1), intended to signal the driver to pay attention to the road immediately. In situations with very high visual demands, such as approaching an intersection or a tight curve, the warning threshold was set to zero. In these situations, the warning icon was displayed immediately based on the GPS and map information, regardless of the driver looking at the phone or not, in order to minimize gaze time on the screen.

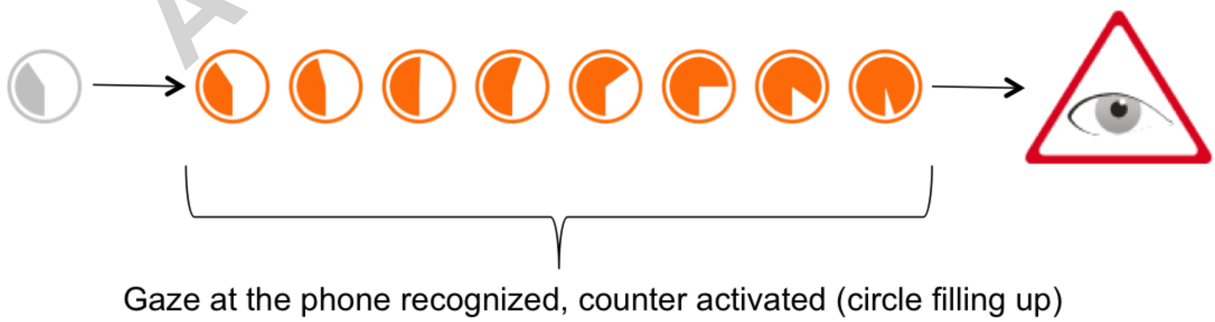

Figure 3. The user interface icons: the circle symbols and the warning icon, and their relative sizes.

The application does not restrict the usage of the smart phone in any way. The idea is to leave the final responsibility of safe driving to the driver, but the application is intended to help the driver in estimating the risks and demands of her or his behavior. The application is not yet available for the public.

4 METHOD

In order to test the effects of the VisGuard prototype on drivers' in-car glancing behaviors and to study how the drivers experience the application, we organized a test track experiment and a survey with 31 participants.

4.1 Design and Hypotheses

The mixed-model experimental design of 2 x 3 x 2 focused on the effects of the following independent variables; Warnings: Control/Warnings (within); In-Car Task Type: Calculator/Navigation/Text message (within); and Driving Experience: Novices/Experienced (between). In particular, we looked at in-car glance durations, their distributions, and the ratio of glance time on road of the total task duration.

The dependent variables were glance time on road (%), percentage of over-2-second in-car glances, and median in-car glance duration. In particular, the experiment tested whether the VisGuard warnings affect these safety-relevant visual behaviors. Glance time on road measures how much visual attention the driver devotes to the driving environment during an in-car task. Low percentages suggest that in-car tasks are performed as fast as possible, without regard for the situational driving conditions. Previously, for instance, Volvo has utilized this metric in testing their distraction warning systems (Wege and Victor, 2014). The ratio of in-car glances over 2 seconds to all in-car glances in percentage was chosen as another metric, because it is a standard verification criterion in test guidelines for in-vehicle electronic devices (NHTSA, 2013b) and it is known to have an association to elevated level of a safety-critical incident risk in real traffic (Liang et al., 2012). In a similar fashion, if the median in-car glance duration is high, it suggests a high number of long safety-critical in-car glances. According to Wierwille's (1993) general visual sampling model, drivers tend to keep off-road glance durations between on average 500 to 1600 ms in most traffic scenarios. Median glance durations for an in-car task above 1600 ms could suggest high visual in-car task demands that interfere with this general visual sampling tendency.

Furthermore, it is known that complexity of the in-car task increases the cognitive demands of the task, which can lead to cognitive capture of attention (e.g., Young and Salmon, 2012; Baumann, Petzoldt, Groenewoud, Hogema, and Krems, 2008; Blanco, Biever, Gallagher, and Dingus, 2006). We studied three different in-car tasks with varying interaction demands and complexity (see Table 1) in order to see if the in-car task demands interact with the warning efficiency.

In addition, we were interested if the warnings are more useful for novice drivers than for the experienced ones, as could be expected from the vast literature on risks related to novice drivers (e.g., Wikman, Nieminen, and Summala, 1998). Experienced drivers are expected to have a better situation awareness than the novice drivers already in the control condition (Underwood, Chapman, Bowden, and Crundall, 2002) as well as to have lower in-car glance durations in general (Wikman et al., 1998). The study by Donmez et al. (2010) suggests that the warnings could be beneficial for high-risk novice drivers, in particular. Based on the factorial design and the relevant literature, we formulated three hypotheses:

- H1. The warnings increase glance time on road and decrease individual in-car glance durations.
- H2. The warnings are more efficient for novice than experienced drivers.
    - The interaction Warnings x Driving experience should be significant and there should be greater increase in glance time on road as well as greater reduction

for Novices on the percentage of over-2-s in-car glances and median in-car glance durations than for Experienced from Control to Warnings conditions.

- H3. The efficiency of the warnings is in-car task-dependent.
  - This means that there should be significant interactions Warnings x In-Car Task Type on the dependent measures. If these are observed, more detailed analyses on the differences between the tasks are conducted.

In addition, the efficiency of the location-based (WT=0) warnings while multitasking will be further analyzed by taking the most demanding part of the test track under a closer analysis. The survey results should indicate the participants' experiences and in particular, the acceptance, of the warning system:

- H4. The participants experience the warnings as acceptable.
  - The means for the constructed experience factors differ to a positive direction from the midpoint of the scale, indicating positive general experiences towards the application.

4.2 Participants

In total 31 participants were recruited via university's student emailing lists. The requirements for the participants included experience on smart phones, a valid driving license, normal or corrected vision, and either less than 2,000 km (novice drivers) or more than 50,000 km (experienced drivers) of lifetime driving experience. Twenty of the participants were men (64.5%) and 11 were women (35.5%). The ages of the participants varied between 18 and 67 years with the mean age being 31.2 years (SD 12.1). The participants had had their driving license on average 12.3 years (SD 11.6). There were 10 novice (in or just out of driving school with less than 2,000 km of lifetime driving experience) and 21 experienced (with over 50,000 km of lifetime driving experience) drivers. Due to technical difficulties with the face recognition during the practice trials, one novice and one experienced driver did not drive the trials, but they were introduced to the system with a demonstration and they filled in the survey based on this experience (N=31 for the survey). Due to vibrations of the car and low levels of light during trials driven after sunset, the quality of the video image of these trials proved to be challenging to code the eye-movements reliably afterwards from the videos. This led to a sample size of N=24 (9 novices and 15 experienced drivers) for the glance analyses. All the participants were rewarded with a movie ticket, a mobile car holder, and a car charger.

4.3 Apparatus

The experiment was conducted on a driving practice track in Lievestuore, Finland during two consecutive weekends in early March 2014. The track was closed from other vehicles. The utilized car was a Volkswagen Golf GTI 2005 with direct shift automatic transmission to make the driving task as fluent and easy as possible (Figure 4).

Figure 4. The car on the Lievestuore track.

The driving paths on the track and the highly demanding parts of the track (WT=0) are illustrated in Figure 5. The most demanding parts of the track (circles in Figure 5) included three intersections as well as four tight bends with low visibility. The most demanding path on the track included a three-way intersection after an icy downhill slope, a stop sign and a tight (over 90 degrees) turn to right with low visibility. As an additional “warning sign”, there was a wrecked car on the right-hand side of the road, which marked the start of the WT=0 area in the intersection. The driving paths and directions were varied between the trials and participants in order to mitigate unwanted learning effects.

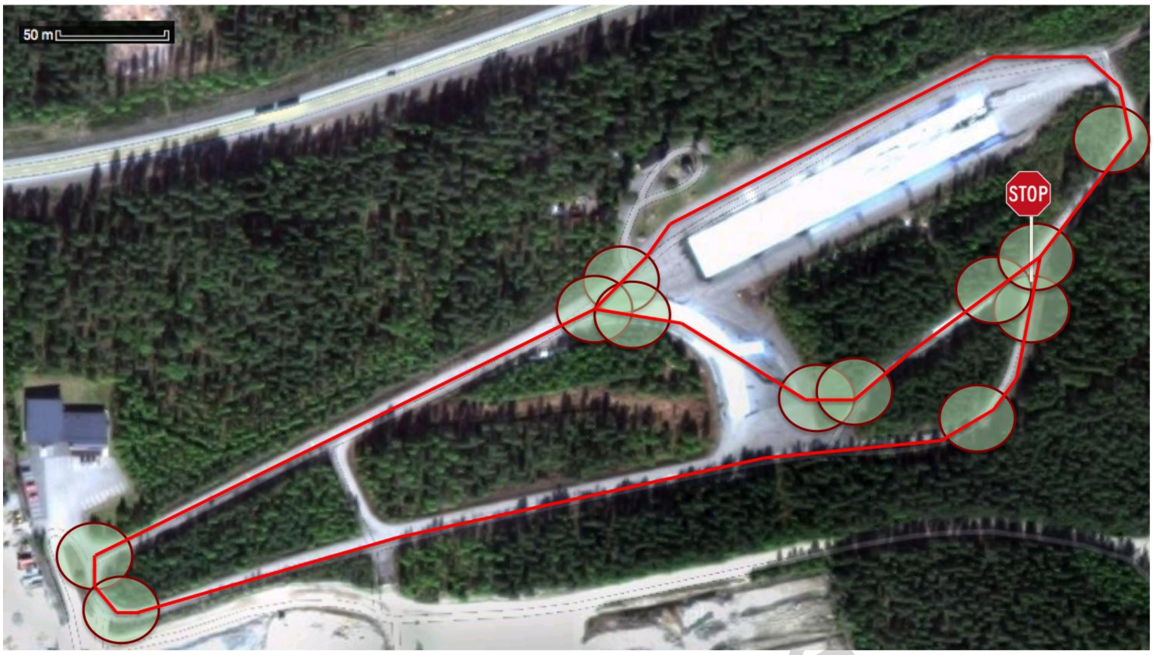

Figure 5. The driving paths on the track. The circles indicate high demand areas with WT=0, the intersection marked with stop a sign being the most demanding area.

Two Samsung Galaxy S3 smart phones (GT-I9305, Android 4.4.2) had the VisGuard application installed and these were alternated between the Warnings blocks. The dashboard holder was securely installed on the passenger side air vent, as seen in Figure 2. The angle of the phone towards the driver was adjustable, and was adjusted for each driver for optimal gaze recognition before each trial. In the other phone, the VisGuard warnings were set off, and in the other, they were set on. The phones had differing content (e.g., text messages), but the search targets were always set on the same locations. In addition, the smart phone with the warnings on was switched between days. These controls kept task complexity at a similar level across the conditions.

In contrast to possible real-world application settings, in the experiment, the same warning times for both novice and experienced drivers were used in order to avoid this confounding variable on the analyses of the effects of driving experience. Because of the fairly static and low visual demands of the practice driving track (excluding the areas with WT=0) and the low, but fairly constant driving speed (max 40 km/h), the WTs remained at almost static ca. 1.7 s levels on the relatively straight parts of the track.

A Sony HD video camera was used to capture driver's gaze and the screen of the smart phone in the holder. The camera was on a holder attached to the backseat of the car and operated by a research assistant who was recording the driver's gaze via a rearview mirror. An extra-large rearview mirror placed above the standard rearview mirror was used in order to keep the driver's eye constantly visible in the video image. Other equipment included a laptop for filling in the two (pre- and post-experiment) questionnaires online.

4.4 Procedure

At the arrival to the driving track, the participants filled in a consent form and a pre-study questionnaire for demographic data. After this phase, they completed four practice tasks with a Galaxy S3 smart phone while being stationary. The conducted tasks were similar to the tasks used later in the experiment. The participants were given the time they needed to get comfortable with driving on the track and the car with its controls. Two practice tasks without warnings were performed while driving; a task of using the calculator to count "89-56" and a task of searching and setting a destination (by typing and selecting a suggestion) to Tampere. Before the experiment and between each trial, a gaze recognition test was performed and the car phone holder adjusted, if needed.

In the trials, the in-car tasks started always on the home screen and the car was stationary at the starting position in the big crossings (in the middle of Figure 5) before the participant said to be ready to start the next trial. The participants were briefed to focus on the driving, to try to keep the speed near the assigned speed limit of 40 km/h if possible, and to try to keep the car on its lane. However, the participants were asked to adjust the speed according to the situation and their own feeling. In addition, the participants were instructed to drive as if they are driving in real traffic and to obey traffic regulations. For safety reasons, there was no other traffic on the track during the trials. However, as there was no fence on the track, there was always a possibility of other road users. Therefore, the participants were instructed to pay attention to the road. Even though the safety measures were strictly enforced, in three trials the risk of other road users was realized when hikers were crossing the track. However, no safety-critical incidents occurred during these events.

The participants were further instructed that there is always the possibility of unexpected events, as for example the road surface can be slippery. Therefore, they were instructed to take their time in completing the tasks with the smart phone. They were also told to use their own discretion in completing the tasks. However, they were instructed to keep the car in movement, if possible, when completing the smart phone tasks. They were instructed to take turns only according to the verbal instructions given by the experimenter sitting on the passenger seat. An important last advice was that if they would hear someone saying STOP, they should brake the car to a halt immediately.

The in-car tasks were selected to be equal with realistic in-car activities one could imagine drivers are willing to engage in while driving (see Table 1). The in-car tasks varied in many aspects, but the Gallery task was used as a practice task to reduce unwanted learning effects. It was the simplest task at least in terms of the number of interaction steps required as well as the number of possible interaction options per screen. The last Text message task was intended to be the most visually complex one, as it involved visual search of finding five targets in an unordered list among 13 items in total in a sentence-like text message. Blanco et al. (2006) as well as Kujala and Saariluoma (2011) have shown that finding a semantic target on a compressed text is a visually highly demanding task while driving. In addition, it included a small cognitive decision-making component of choosing if an item is a member of the target category (dairy product or fruit), whereas the other tasks included only a memory component of remembering the verbal target.

Table 1

In-Car Tasks (Trials) in the Order of Execution per Block (Top-Down).

Task type	Black phone	White phone	Task steps
Gallery	Find an image of a dog (with page-by-page scrolling)	Find an image of a sheep (with page-by-page scrolling)	1. Gallery 2. Browse (7 items)
Calculator	Use the calculator to count 851/742, and 269*358	Use the calculator to count 962*853, and 158/247	1. Calculator (on the Home screen) 2. (E.g.) 962*853= 3. (E.g.) 158/247=
Navigation	Search and set a destination to Helsinki, after which stop the navigation and search and set a destination to Oulu	Search and set a destination to Jyväskylä, after which stop the navigation and search and set a destination to Turku	1. Google Maps (on the Home screen) 2. Search 3. (E.g.) h 4. (E.g.) Helsinki 5. Car icon 6. Select a route 7. Start 8. Stop the guidance (and go to 2.)
Text message	Read a text message – read aloud what dairy products you have to buy from the grocery store	Read a text message – read aloud what fruits you have to buy from the grocery store	1. Messages (on the Home screen) 2. Select the first message in the menu 3. Read the targets aloud (13 items with five targets)

The tasks with the same phone (i.e., warnings off or on) were performed in a sequence in a block, but the orders of the Control and Warnings blocks were counter-balanced across the sample and level of driving experience. In addition, the exact contents of the task varied with the phone (black or white, see Table 1) and therefore, from the Control to Warnings blocks from day to day.

After the trials, the participants filled in a post-study questionnaire on their experiences, after which they were rewarded. A single experiment lasted on average one hour and 15 minutes.

4.5 Survey Design

Before the experiment, the participants filled in a web-based pre-study questionnaire, which included questions about the participants' background such as gender, age, and amount of years the participants had had a driving license. The results of the pre-study questionnaire were mostly used in analyzing the demographic data from the participants, which is presented in the "4.2 Participants" section of this paper.

After the experiment, the participants filled in a web-based post-study questionnaire, which included statements (i.e., questions or items) about the VisGuard application. Most of these statements were positive in nature (e.g., "The application was useful for me"), but the questionnaire included also some negative statements (e.g., "The application was annoying"). The answering options for the statements were in a 5-point rating scale with response options ranging from "strongly disagree" (1) to "strongly agree" (5). Response option 3 was "neither agree nor disagree". Each statement had also an additional text field for possible open comments regarding the chosen answering option. The post-study questionnaire items and the constructed factors based on these are presented in the "4.6 Analysis" section of this paper.

In detail, the post-study questionnaire included questions about different themes, which were based on previous literature. From research regarding trust in technology, the questionnaire had three items related to the application's trustworthiness (Lee and See, 2004; Dzindolet et al, 2003), perceived consistency (i.e., reliability) (Lee and See, 2004; Bisantz and Seong, 2001), and timeliness (Grandison and Sloman, 2000). In addition, two items were related to whether the participant would recommend the application (Jonsson, Harris and Clifford, 2008; Ghazizadeh, Lee and Boyle, 2012) and to application designers' benevolence (Lee and See, 2004; Mayer, Davis and Schoorman, 1995).

From technology acceptance research, two items were related to the application's and its warnings' perceived usefulness (Roberts, Ghazizadeh, and Lee, 2012; Pavlou, 2003;

Venkatesh and Davis, 2000) and another two on whether the participant would use the application after the study (Venkatesh and Davis, 2000). Also, four items were related to the general acceptance of the application and user satisfaction (Pavlou 2003; Venkatesh and Davis, 2000; Lewis, 1995; Van der Laan, Heino and De Waard, 1997). In contrast to perceived usefulness, four items were related to the perceived harmfulness (Bisantz and Seong, 2001) and annoyance (Van der Laan, Heino and De Waard, 1997; Weinstock, Oron-Gilad and Parmet, 2012) related to the application.

In addition, four items were related to the suitability of the application for its intended task (i.e., validity) (Lee and See, 2004; Venkatesh and Davis, 2000). The application developers' definition of the application's intended task was defined before these statements in the questionnaire. Finally, further four items related to the participants' experiences with the functioning of the circle symbol of the application. The questionnaire included also other items regarding, for example, the participants' opinions about the commercialization of the application, but the results of those items are not analyzed or reported here as they are not in the focus of this paper.

4.6 Analysis

Because of the expected unreliability in the glance data collected with VisGuard, two data reducers coded independently the in-car glance durations to the smart phone from video after the SAE-J2396 standard (SAE, 2000). Noldus Observer XT software was used for coding the glance data. Inter-rater reliability was assessed by calculating Cohen's Kappa and Intraclass Correlation Coefficient (ICC) for the glances of eight randomly selected trials the both data reducers had scored. Both metrics indicated high levels of inter-rater reliability (Kappa=0.874, 95% CIs [0.848, 0.900], N=160; ICC=.910, 95% CIs [.871, .937], N=122). For Kappa, there were 140 events during which both data reducers scored a glance within 500 ms of each other, and a total of 20 events when only one of them did.

Repeated measures ANOVA and paired samples t-tests were used for testing the hypotheses on the glance metrics. For each ANOVA, assumptions of sphericity were confirmed. If the assumption of sphericity was violated, degrees of freedom were adjusted with the Greenhouse-Geisser correction.

The received questionnaire answers were analyzed with exploratory factor analysis (Principal Axis Factoring) using IBM SPSS Statistics (Version 20). The used rotation method was Varimax with Kaiser Normalization. The purpose of the exploratory factor analysis was not to create a novel general scale but to reduce the data set by constructing the most suitable factors amongst the items for the purposes of the current experiment (H4). Twenty-five questionnaire items that were based on the previous research presented in “4.5 Survey Design”, and which correlated at least with $r = .5$ with at least one other item were selected to the initial factor analysis.

The factor analysis was done five times, because some of the items originally thought to contribute for the acceptance-related factors had to be excluded from further analysis. On the first run, "I could recommend this application to my friends" had high cross loadings (> .40), "The intentions of the application’s designers are good" had no loadings at all, and "I am satisfied with the application" had high cross loadings on the second run. On the third run, the items "The application supported my driving performance", "The application supports my safe driving", "The warnings support my safe driving", and "The circle symbol did not have a harmful effect on my driving performance", had still high cross-loadings, and were removed. On the fourth run, items "The warnings given by the application were annoying" and "The application increased my alertness in traffic" had high cross-loadings, and one of the factors had only two items with loadings over .40; "I was happy to use the application" and "I accepted the application as part of my driving activity during the study". All these items were removed before the fifth run. In addition, the item "The application works well in its intended task" was removed from the final solution, as it loaded only for a factor otherwise closely related to the circle symbol.

For the final solution with 13 items, the Kaiser-Meyer-Olkin measure of sampling adequacy was .65, that is, greater than the limit of acceptable (.50, Kaiser, 1974), and Bartlett's test of sphericity indicated high significance, $\chi^2(78) = 250.643$, $p<.001$, indicating that factor analysis is appropriate. All the communalities were over .40. Following the Kaiser criterion, only factors with Eigenvalues over 1.0 were selected. This resulted in the selection of four factors. Scree-plot indicated the last notable drop in the Eigenvalues after the fourth factor, after which there was a less steep decline. The rotation converged in 5 iterations. The factor loadings are presented in Table 2.

Table 2

Factor Loadings for Exploratory Factor Analysis with Varimax Rotation of the Questionnaire Items.

Item	Factor 1	Factor 2	Factor 3	Factor 4
The warnings given by the application appear when they are needed.	.76	.33	-.27	.12
The application works coherently and logically.	.62	.03	-.29	.21
The application is trustworthy.	.63	.39	.05	.29
After this study, I could use the application every day while driving.	.33	.79	-.21	.14
I could use this application after the test run.	.11	.73	-.14	.23
The warnings that the application gives are useful for me.	.24	.80	-.16	-.04
The application was useful for me.	.02	.78	-.10	-.09
The warnings had a harmful effect on my driving performance.	-.07	-.28	.78	-.01
The application had a harmful effect on my driving performance.	-.26	-.05	.78	-.14
The application was annoying.	-.12	-.18	.77	-.29
The circle symbol was not annoying. [1]	.02	-.08	-.12	.78
The circle symbol supports my safe driving.	.29	.17	-.25	.76
The circle symbol is useful for me.	.23	.09	-.05	.69

Notes [1]Reverse-coded values. Factor loadings > .40 are in boldface. N=31.

Factors were constructed from items that had a loading of over .40, but only if the Cronbach's alpha was over .70. The factors were finally selected and labeled according to themes identified based on the literature presented in "4.5 Survey Design". Table 3 presents the questionnaire items that contributed to the four different factors identified in the factor analysis. All the four factors can be interpreted as indicative of an underlying factor contributing to the acceptance of the application. The four factors explained 67.1% of the total variance of the items. Finally, means for the constructed factors were calculated by adding up the scores of each item in a factor and dividing the total by the number of items included in the factor. We hypothesized that the scale means for the constructed factors differed to a positive direction from the theoretical mean (or median) of 3.0 (one-sample t-test and one-sample Wilcoxon Signed Rank test), indicating a positive general experience on the latent factors (see H4).

Table 3

Factor Loading of Each Questionnaire Item and Cronbach's Alpha for Each Identified Factor.

Factor number and short label / Item (Factor long label) (% of total variance)	Factor Loading	α
Factor 1: Trust (Trust in the application) (13.4%)[1]		.79
The warnings given by the application appear when they are needed	.76	
The application works coherently and logically	.62	
The application is trustworthy	.63	
Factor 2: Usefulness (Usefulness of the application) (21.8%)[1]		.88
After this study, I could use the application every day while driving	.79	
I could use this application after the test run	.73	
The warnings that the application gives are useful for me	.80	
The application was useful for me	.78	
Factor 3: Harmfulness (Harmfulness / annoyingness of the application) (16.5%)[1]		.85
The warnings had a harmful effect on my driving performance	.78	
The application had a harmful effect on my driving performance	.78	
The application was annoying	.77	
Factor 4: Circle symbol (Functioning of the circle symbol) (15.4%)[1]		.82
The circle symbol was not annoying[2]	.78	
The circle symbol supports my safe driving	.69	
The circle symbol is useful for me	.78	

Notes [1]Rotation sum of squared loading (% of variance), [2] Reverse-coded values. N=31.

For all the statistical analyses the alpha level was set to .05. For multiple comparisons, Bonferroni correction was applied for the alpha level. Where applicable, partial eta-squared and Cohen's d were used as a measure of effect size.

5 RESULTS AND DISCUSSION

5.1 Glance Metrics

The glance metrics indicated varying levels of support for the hypotheses. The glance metrics relevant for the three hypotheses H1-H3 are presented in Table 4.

Table 4

Glance Metrics Relevant for the Hypotheses H1-H3 (N=24). Values are Means (Standard Error of Mean).

	Glance time on road (%)		Percentage of over-2-s in-car glances (%)		Median in-car glance duration (s)	
	Control	Warnings	Control	Warnings	Control	Warnings
Total	55.4 (2.7)	62.3 (2.3)	28.1 (4.1)	29.1 (4.4)	1.57 (.09)	1.58 (.09)
Calculator	60.5 (2.6)	63.7 (2.5)	24.9 (4.4)	22.4 (4.0)	1.49 (.09)	1.42 (.08)
Navigation	58.6 (2.8)	61.6 (2.6)	24.1 (3.7)	26.6 (4.2)	1.44 (.09)	1.52 (.09)
Text message	47.0 (3.6)	61.0 (2.7)	30.6 (4.7)	34.0 (5.7)	1.70 (.10)	1.75 (.11)
Novices	56.9 (4.2)	62.7 (3.6)	32.6 (6.3)	31.0 (6.7)	1.66 (.13)	1.59 (.14)
Experienced	53.9 (3.5)	61.8 (3.0)	23.5 (5.3)	27.3 (5.6)	1.49 (.11)	1.57 (.11)

A significant main effect of the VisGuard warnings was observed on glance time on road, which was increased on average by 12.5 percent, $F(1,20)=13.125$, $p=.002$ partial $\eta^2=.396$. However, no significant effects were found on percentage of over-2-second in-car glances ($p=.743$) or on median in-car glance durations ($p=.908$). Therefore, Hypothesis 1 was supported by the fact that the warnings increased glance time on road, but not supported by the absent effects on individual in-car glance durations.

There were no significant interaction effects of warnings and driving experience on glance time on road ($p=.578$), percentage of over-2-second in-car glances ($p=.409$), or median in-car glance duration ($p=.301$). This finding suggests the effects of the warnings are independent of the driving experience of the driver (i.e., Hypothesis 2 is rejected). In general, the novice drivers seemed to have a greater number of over-2-second in-car glances than the experienced (see Table 4), but the difference was not significant with this sample size (only 9 novice drivers). Due to the unequal group sizes, we tested the main effect of driving experience on in-car glance metrics in the control and experiment conditions by averaging over the tasks, and by testing the differences also with Welch's t-test that is more reliable when the two samples have unequal sample sizes. Still, we did not find significant differences between the driving experience groups with any of the in-car glance metrics (glance time on road, control: $p=.566$, experiment: $p=.843$; percentage of over-2-second in-car glances, control: $p=.274$, experiment: $p=.670$; median in-car glance duration, control: $p=.308$, experiment: $p=.904$).

There was a significant interaction effect of Warnings and In-Car Task Type on glance time on road ($F(2,28.198)=10.975$, $p=.001$, partial $\eta^2=.354$), but not on percentage of over-2-second in-car glances ($p=.579$), or on median in-car glance durations ($p=.434$). These findings give partial support for Hypothesis 3; the efficiency of the warnings on glance time on road seems to be in-car task-dependent. However, more detailed analyses are required in order to better understand the task type effects on the efficiency of the warnings.

All the in-car task types proved to be visually demanding because the mean percentages of over-2-second in-car glances were well above the 15 percent for the 85th percentile verification criteria for in-car tasks set by NHTSA (2013b) (Calculator: 25.7 [SEM=4.0], Navigation: 27.5 [SEM=3.9], Text Message: 32.6 [SEM=4.8]). The driving speeds were low (40 km/h) compared to the NHTSA (2013b) testing scenario (80 km/h), which may partly explain the high percentages. On the other hand, the visual sampling model of Wierwille (1993) suggests that drivers prefer to keep in-car glance durations well below 2 seconds in most traffic situations.

More detailed analysis revealed that there were significant differences between the in-car task types on glance time on road ($F(2,40)=7.485$, $p=.002$, partial $\eta^2=.272$), percentage of over-2-second in-car glances ($F(2,40)=3.333$, $p=.046$, partial $\eta^2=.143$), as well as median in-car glance durations ($F(2,40)=11.800$, $p<.001$, partial $\eta^2=.371$). The mean differences between the in-car task types are displayed in Table 5. Paired comparisons revealed, as expected, that the Text message task seemed to be the most visually complex task with the glance time on road as well as on the median in-car glance duration metrics. Besides the Text message task, the participants seemed to be able to follow the general visual sampling behavior suggested by Wierwille (1993); keeping the median in-car glance durations below 1.6 seconds (see Table 4).

Table 5

Mean Differences Between In-Car Task Types (N=24).

	Glance time on road (%)	Percentage of over-2-s in-car glances (%)	Median in-car glance duration (s)
Text message - Calculator	-8.9** (p=.001)	7.0 (p=.024)	.243*** (p<.001)
Text message - Navigation	-6.0 (p=.032)	5.2 (p=.051)	.204** (p=.001)
Navigation - Calculator	-2.8 (p=.200)	1.8 (p=.562)	.039 (p=.518)

Notes * p < .017 (Bonferrroni-adjusted alpha level for multiple comparisons), ** p < .01, *** p <.001

However, the greatest effect of the warnings on glance time on road was also seen in the visually most complex Text message task (see Figure 6). The percentual increases on glance time on road from Control to Warnings condition per task type were 5.3% for the Calculator ($t(23)=1.393$, $p=.177$), 5.1% for the Navigation ($t(23)=2.210$, $p=.037$, $d=.225$), and 29.7% for the Text message ($t(23)=4.407$, $p<.001$, $d=.929$). The increases on glance time on road were achieved without significant increases on individual in-car glance durations (Table 4). Because of confounding factors, we cannot say definitely that the visual complexity of the in-car task affected the efficiency of the warnings but it seems to be one plausible factor.

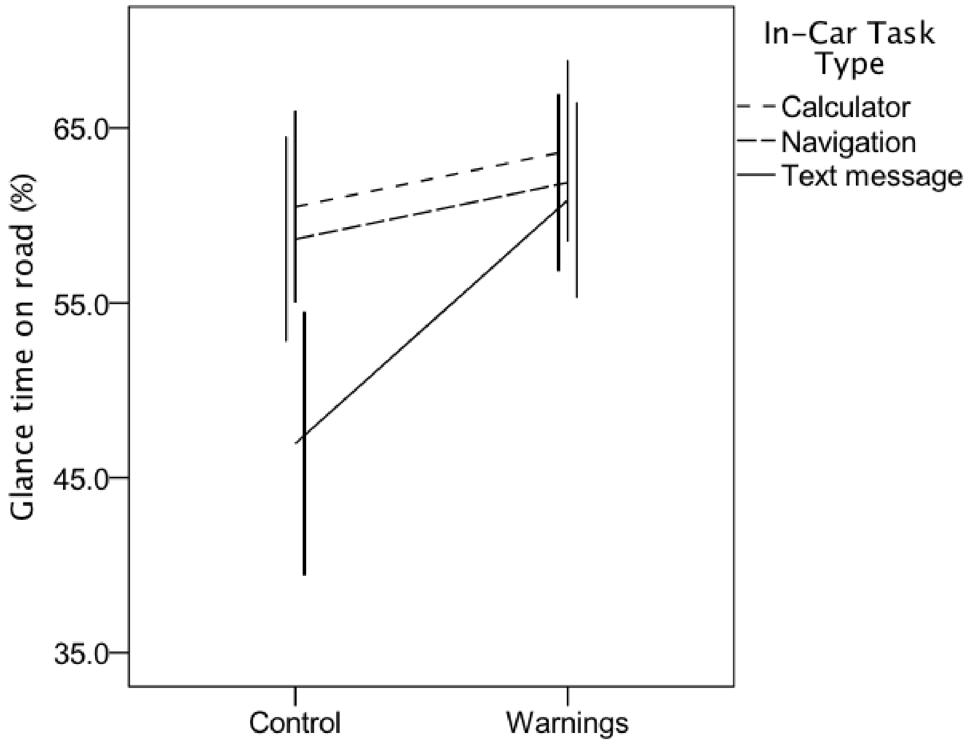

Figure 6. Mean glance time on road (%) by Warnings and In-Car Task Type (N=24). The bars represent 95% CIs.

Due to varying lighting conditions, we noted a number of technical difficulties in gaze tracking with the VisGuard application, which may at least partly explain the absent effects of the warnings on the individual in-car glance durations. Therefore, we wanted to see if the observed effects of warnings on glance time on road would be due to the location-based warnings (WT=0) that worked reliably based on the GPS signal, and thus, did not require the gaze tracking. We took the glance time on road in the most demanding intersection on the track with the stop sign (see Figure 5) and WT=0 under closer analysis. The driving directions and paths were systematically varied between the participants in order to mitigate unwanted order effects, but 14 participants drove the Calculator and Text message trials on exactly the same path that included the stop sign intersection for both Control and Warning conditions. The sample of the 14 participants was in balance for the trial orders. This enabled us to analyse if the general task type effect was visible in the glance time on road also in this highly demanding intersection. The glance time on road was coded frame-by-frame from the video material from a clearly visible landmark (the wrecked car on the side of the road) leaving the picture as marking the starting point of the WT=0 section to another clear landmark (a ploughing stick) leaving the picture as marking the end point of the section identically for both Control and Warnings conditions.

Figure 7 illustrates that, at a general level, the participants seemed to acknowledge that the intersection required more visual attention than driving in general and in the Calculator trials they were successful in devoting high levels of visual attention on the road environment. However, in the Text message trial in the Control condition, the mean glance time on road was only 71.3% compared to the 93.5% in the Calculator trial in the Control condition ($t(13)=2.971$, $p=.011$, $d=1.105$). The warnings succeeded in raising the percentage up to 90.1% (26.4% increase) in the Text message Warnings trials ($t(13)=2.761$, $p=.016$, $d=.838$. These findings give further support for the result that the Text message tasks were the most engaging ones, capturing participants' attention also in highly demanding situations and perhaps undermining their situation awareness. However, the location-based warnings (WT=0) seemed to increase participants' attention on road on these tasks, at least on this highly demanding point of the track.

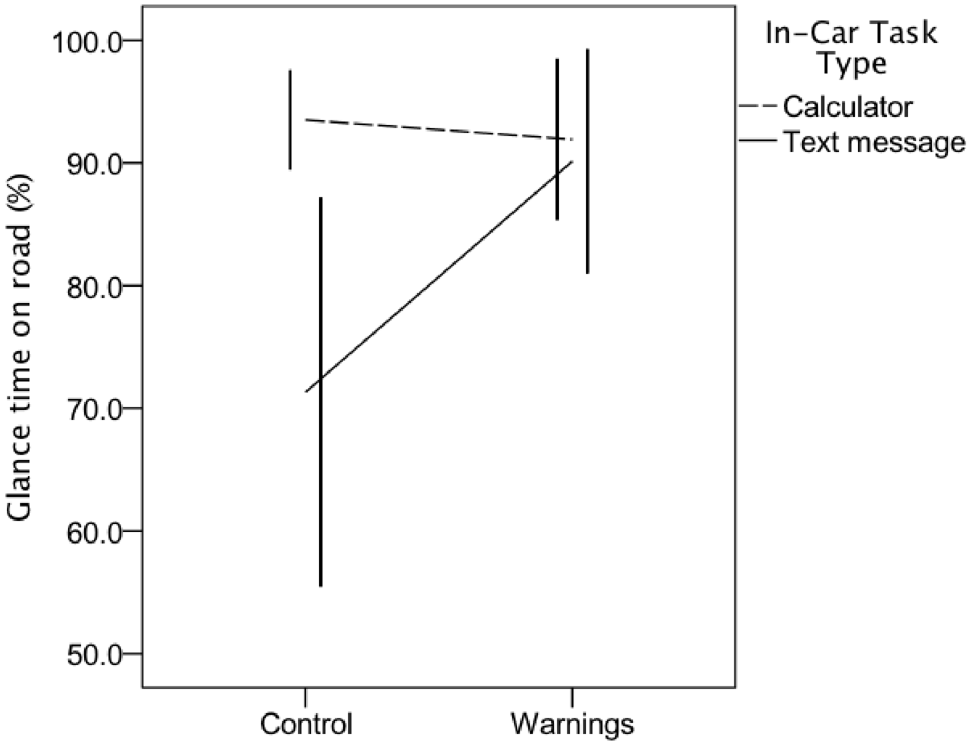

Figure 7. Glance time on road (%) by Warnings and In-Car Task Type in the most demanding intersection on the track with WT=0 (N=14). The bars represent 95% CIs.

Overall, the glance metrics indicated that the warnings increased glance time on road, and in particular in the visually most complex Text message trials, whereas no effects on individual in-car glance durations were found. Driving experience, or the lack of it, did not affect the efficiency of the warnings. The missing effects on in-car glance durations by the warnings can be attributed to the poor working of the gaze tracking in real driving environment with varying lighting conditions. However, the location-based warnings (WT=0) seemed to have a clear impact on the participants' glance time on road and in particular in the visually complex Text message tasks. This finding is well reflected in the analysis on the glance time on road in the most demanding intersection, as illustrated above, but also in the survey results on the participants' experiences towards the application.

5.2 Survey Results

Table 6 presents the means and standard deviations for the four different factors identified in the factor analysis as well as for the individual questionnaire items that contributed to the factors.

Table 6

Mean and Standard Deviation for Each Identified Factor and Contributing Questionnaire Item.

Factor number and short label (Factor long label) / Item	M	SD
Factor 1: Trust (Trust in the application)	3.71	.68
The warnings given by the application appear when they are needed	3.61	.96
The application works coherently and logically	3.94	.68
The application is trustworthy	3.58	.77
Factor 2: Usefulness (Usefulness of the application)	3.45	.91
After this study, I could use the application every day while driving	3.10	1.25
I could use this application after the test run	3.77	.88
The warnings that the application gives are useful for me	3.68	1.11
The application was useful for me	3.26	.97
Factor 3: Harmfulness (Harmfulness and annoyingness of the application)	2.06	.92
The warnings had a harmful effect on my driving performance	1.94	.93
The application had a harmful effect on my driving performance	2.06	1.15
The application was annoying	2.19	1.05
Factor 4: Circle symbol (Functioning of the circle symbol)	3.23	1.02
The circle symbol was not annoying[1]	3.55	1.15
The circle symbol supports my safe driving	3.06	1.21
The circle symbol is useful for me	3.06	1.21

Notes [1] Reverse-coded values. N=31.

As can be seen from Table 6, in general, the means of the identified factors indicated positive experiences of the participants towards the application. The participants' textual questionnaire comments regarding the application also supported the received mean values of the factors. For example, the rather high level of trust in the application (M = 3.71, SD = .68) was reflected in the following participant comment: "The application had relevant warnings in dangerous situations. The detection of dangerous situations works well". The usefulness of the application (M = 3.45, SD = .91) was commended, for example, in the following way: "It is a useful application, which increases the safety of driving".

The low level of experienced harmfulness and annoyingness of the application (M = 2.06, SD = .92) was supported by the following comment: "The usage of the application did not have a harmful effect on my driving performance. Instead, I think it improved it". Finally, Factor 4: Functioning of the circle symbol (M = 3.23, SD = 1.02) received a mean only slightly larger than 3.0. In line with this result, the participants commented the circle symbol to be "rather unnoticeable" and "difficult to read". Furthermore, one participant commented that "The warning triangle works nicely, but the circle symbol does not – it is difficult to perceive".

Hypothesis 4 that the scale means for the constructed factors differ to a positive direction from the theoretical mean of 3.0 was tested with one-sample t-test. In addition, the result of non-parametric one-sample Wilcoxon Signed Rank test (with 95% confidence level) was used to support the one-sample t-test result with each factor. The results of these analyses are described in Table 7 along with the short labels of the factors. Based on these results, the means differ statistically significantly from 3.0 on all the other factors, except on Factor 4 (Circle symbol). The reason for Factor 4 to not differ statistically significantly from 3.0 may be that the participants did not seem to experience the circle symbol to be neither very useful nor harmful. Rather, their attitudes were neutral towards the circle symbol. This was reflected in participant comments such as "I think the circle symbol did not have an effect on my behavior" and "I focused more on the given task and on the driving [than on the circle symbol]".

Table 7

Factors' One-sample Test Mean Differences, One-Sample t(30) values and Wilcoxon Signed Rank Test Z Values With Their Significance Levels.

Factor's short label	One-sample test mean difference	One-sample t(30) value	Wilcoxon Signed Rank Z value
Factor 1: Trust	0.71	5.84^{***}	4.05^{***}
Factor 2: Usefulness	0.45	2.78^{**}	2.28^{*}
Factor 3: Harmfulness	-0.94	-5.68^{***}	-4.05^{***}
Factor 4: Circle symbol	0.23	1.23	1.26

Notes * $p < .05$, ** $p < .01$, *** $p < .001$

With all the other factors (than Factor 4), the means of the factors were statistically significantly greater than 3.0, except with Factor 3 (Harmfulness), which was significantly lower than 3.0. This result is due to the fact that, in contrast to the other factors, all the statements regarding the application in Factor 3 were negative.

Pearson correlations between the factors were calculated (see Table 8) to establish the construct validity of the factors. Table 8 indicates that the correlations between the factors were otherwise statistically significant (at least on the .05 level), except that the factors "Usefulness of the application" or "Harmfulness and annoyingness of the application" did not correlate with the "Functioning of the circle symbol" factor. Based on this result — and also on the comments given by the participants regarding the circle symbol — it can be concluded that the circle symbol's functioning was not experienced to have importance over the general perceived usefulness (or harmfulness) of the application. Rather, especially the warnings given by the application were experienced to contribute to the perceived usefulness of the application, which became also evident in participant comments like the following one: "Yes, the warning triangle was useful, and it could have been even more visible".

Table 8

Correlation Matrix Presenting the Values (Pearson's r) for Inter-Correlations between the Identified Factors.

Factor short label	Trust	Usefulness	Harmfulness	Circle symbol
Trust	1			-
Usefulness	0.52^{**}	1		
Harmfulness	-0.42^{*}	-0.38^{*}	1	
Circle symbol	0.43^{*}	0.20	-0.35	1

Notes * $p < .05$, ** $p < .01$

There were also other relevant items in the post-study questionnaire than the ones presented in Table 6. These individual questionnaire items are presented in Table 9 in a descending order by the mean value. The individual item means are well in line with the factor means, further supporting the notion that the application was well accepted by the drivers.

Table 9

Mean and Standard Deviation for Each Individual Questionnaire Item Left Out of the Factor Analysis. Sorted by Mean.

Item	M	SD
The intentions of the application's designers are good	4.58	.56
I accepted the application as part of my driving activity during the study	4.19	.70
The warnings support my safe driving	4.06	.77
I could recommend this application to my friends	4.03	.88
I was happy to use the application	3.94	.81
The application supports my safe driving	3.87	1.02
The application works well in its intended task	3.74	.82
The application increased my alertness in traffic	3.74	.82
I am satisfied with the application	3.55	.85
The application supported my driving performance	3.48	.96
The circle symbol did not have a harmful effect on my driving performance[1]	3.42	1.09
The warnings given by the application were annoying	2.16	1.04

Notes [1] Reverse-coded values. N=31.

The four individual items intended to measure the suitability of the application for its intended purpose got all median scores higher than 3.0 (one-sample Wilcoxon Signed Rank test); "The application works well in its intended task" ($Z = 3.77$, $p < .001$), "The application increased my alertness in traffic" ($Z = 3.77$, $p < .001$), "The application supports my safe driving" ($Z = 3.64$, $p < .001$), and "The warnings support my safe driving" ($Z = 4.44$, $p < .001$). The suitability was commented, for example, in the following way: "If you use a mobile (phone) while driving, the application will increase the safety of driving".

Based on the survey results, it can be said on a general level that the application was well accepted by the participants (H4 supported). The application was trusted by the participants as they felt that it works coherently and logically especially in demanding road locations (e.g., tight curves or intersections). The warnings of the application were experienced to be useful and the participants thought they could use the application also after the study. This attitude was also indicated in the survey results regarding the harmfulness and annoyingness of the application and its warning messages, which were experienced to be fairly low. The suitability of the application for its intended task (to support safe driving) was thought to be high, especially in demanding road situations according to the participants. However, the survey results suggest that the functioning of the circle symbol needs reconsideration in order for the application to support safe and distraction-free driving even better.

6 GENERAL DISCUSSION

In this paper, we have studied the effects of a context-sensitive distraction warning application on in-car glance behaviors and subjective experiences of drivers who conducted smart phone tasks on a test track while driving. We have also studied the moderating effects of driving experience and in-car task type on the efficiency of the warnings. The general goal of the study was to better understand the effects of the proactive and context-sensitive distraction warnings on drivers' visual behaviors and acceptance.

The glance metrics indicated a significant increasing effect of the VisGuard warnings on the glance time on road while multitasking with the smart phone. The average increase was 12.5 percent of the total in-car task time. Volvo has reported even 37 percent increases on glance time on road with their Visual Distraction Alert (VDA) systems (Wege and Victor, 2014). However, Volvo's studies have been conducted in driving simulators in highly controlled settings, whereas our experiment was done in a real car with varying environmental conditions, although on a closed track. Our experiences clearly indicated that gaze tracking with current smart phones outside the laboratory is highly unreliable due to varying lighting conditions. Besides the variable environmental lighting conditions, we found the gaze tracking based on face recognition, head pose, and phone holder position in the passenger side air vent problematic. The method requires that the driver turns his/her head sufficiently towards the phone for optimal gaze recognition. In addition, the holder position makes it challenging to utilize peripheral vision for driving while looking at the phone.

The in-car task type had a clear effect on the efficiency of the warnings on the glance time on road. The maximum increase of 29.7% on glance time on road was observed with the visually most complex Text message tasks. The findings are in line with previous research, suggesting that the increased demands of in-car tasks decrease drivers' ability to appropriately evaluate the dynamic visual demands of the driving situation (Young and Salmon, 2012; Baumann et al., 2008; Blanco et al., 2006). According to the data, the context-sensitive distraction warnings can help drivers in overcoming this inability and to place more attention on the road.

However, the in-car tasks differed also in many other aspects than the measured visual demand and these confounding effects should be considered when drawing conclusions. For instance, the Text message reading task included the lowest number of manual interactions (only two inputs to open the message) and the warning icon was displayed always closer to the visual targets (i.e., the message) than in the other tasks. The other tasks' input elements, in particular, were often located on the lower part of the smart phone display (N.B. the icon did not cover the targets in any task). Due to the latter aspect, the warning icon could have been noticed more easily in the Text message trials than in the others. The Text message trials were also always the last trials within a block. However, the data suggests that the visual complexity was significantly higher and the glance time on road considerably lower for the Text message tasks in the Control condition (at least for the most demanding section of the track) than for the other task types. It seems the participants were able to prioritize the driving task at a sufficient level in the Control conditions in the other trials, but not in the Text message trial. The warnings seemed to increase the glance time on road to a comparable level for the Text message task compared to the other tasks. Therefore, it is hard to see how the other (listed) task properties than the visual or cognitive complexity could explain the observed interaction between the warnings and the in-car task type on the glance time on road. The text message task represented the only in-car task in which there was a clear cognitive component together with the brief off-road glances, an unfavorable combination from the viewpoint of missing safety-critical events in the road environment (Lee et al., 2007).

Against our expectations, there were no significant interaction effects of warnings and driving experience on any of the glance metrics. The finding suggests the effects of the warnings are independent of the driving experience of the driver. In general, the novice drivers seemed to have a greater number of risky over-2-second in-car glances in general than the experienced drivers (as in Wikman et al., 1998), but the difference did not become significant with our sample size. In the study by Donmez et al. (2010) their distraction feedback had a significant effect on the glance behavior of the high-risk drivers only. The study suggests there are individual differences in visual sampling behaviors among young, inexperienced drivers, in a similar fashion as one can expect differences among more experienced drivers. Therefore, the novice drivers in our study should probably not be regarded as a homogenous group. However, we did not analyse their glancing behaviors in order to form subgroups due to the small group size. In addition, it should be noted that even if our data does not support differences between novice and experienced drivers, the group sizes were small and unequal, and this finding should be interpreted with caution.

To summarize, the glance metrics indicated that the warnings significantly increased glance time on road, and in particular in the visually most complex Text message trials, whereas no effects on individual in-car glance durations were found. The missing effects on in-car glance durations can be attributed to the poor working of the gaze tracking in a real driving environment with varying lighting conditions. On the other hand, the location-based warnings on the demanding parts of the track (curves, intersections) ahead seemed to have a clear impact on the participants' glance time on road and in the Text message tasks in particular. This finding is well reflected in the analysis of the Text message trials in the most demanding intersection, where glance time on road increased by 26.4% due to the warnings (from 71.3% to 90.1%). The level of driving experience did not seem to affect the efficiency of the warnings.

The survey results indicated a rather high level of trust in the application. This can be a positive result, if the trust is at an appropriate level (i.e., matches the capability of the system), as discussed, for instance, by Lee and See (2004). In addition, the results indicate a high level of perceived usefulness, and low level of perceived harmfulness of the application. All of these factors can be seen to contribute to the general acceptability of the application according to the previous literature as the results of our factor analysis are in line with the Technology Acceptance Model (TAM) and its extensions. From TAM (Venkatesh and Davis, 1996), perceived usefulness (PU) was evaluated especially with the items in "Factor 2:

Usefulness of the application" (Table 6). In our survey, it was not reasonable to measure TAM's perceived ease of use (PEOU), as the VisGuard application did not have or require user input. From a TAM extension labelled Automation Acceptance Model (AAM) by Ghazizadeh et al. (2012), AAM's "trust" is similar to items in our "Factor 1: Trust in the application". In addition, the high mean scores on the individual items on suitability, such as "The application works well in its intended task" and "The application increased my alertness in traffic", are well in line with the factor means and support the importance of the suitability of the application for its intended purpose for technology acceptance (Ghazizadeh et al., 2012). Therefore, we see that the results of our study contribute to the current discussion of what factors affect technology acceptance in general and the acceptance of distraction warning systems in particular (Roberts, Ghazizadeh, and Lee, 2012). However, the interpretation of the results of the exploratory factor analysis should be done with care, as the ratio of cases per items (31 / 13) was low.

Drivers' experiences towards the warning system were significantly more positive than towards the real-time feedback system studied by Roberts, Ghazizadeh, and Lee (2012). We suggest at least three plausible reasons for this finding; 1) the proactive context-sensitivity of the warnings, 2) people in general seem to acknowledge that the use of mobile devices while driving is always a distraction (Jääskeläinen and Pöysti, 2014), and 3) the warnings on the screen of the smart phone were more subtle than the flashing LED + auditory warnings in Roberts, Ghazizadeh, and Lee (2012).

The functioning of the circle symbol, intended to display the remaining warning time threshold, was not thought to be useful by the participants. One plausible explanation for these experiences is the larger than expected delays in gaze tracking. The individual comments also suggest that the circle caused only additional visual load. The warnings, on the other hand, were experienced as highly useful, which suggests that the circle symbols could be removed and the warnings kept as the only icon for guidance. In this study, we expected challenges with the gaze tracking, and therefore, we did not utilize warning sounds, vibrations, or blinking in order to avoid attracting (instead of detracting) the attention of the driver towards the phone in a high-demand situation. However, the effects of other modalities than visual only for the efficiency and acceptability of the warnings should be studied (Smith, Clegg, Heggestad, and Hopp-Levine, 2009).

Existing and suggested distraction warning systems have typically high false positive rates, which can undermine the acceptability of these systems by the drivers (Lee et al., 2013; NHTSA, 2013a). The false positive rate of the VisGuard system was low in the experiment due to the larger than expected recognition delays in the gaze tracking. Consequently, this must have also led to lower true positive rate than intended for the single glance duration warnings (WT~1.7 s), which may partly explain the absent effects on the in-car glance durations. For the location-based warnings of demanding road conditions ahead (WT=0), however, the true positive rate was 100%, as the warnings were displayed based reliably on the GPS signal and map data, regardless of whether the driver's gaze was recognized or not.

Our study had also some shortcomings that should be taken into account in further research. The experience of the participants with the application was fairly short. The acceptability of the application in daily use as well as possibility of long-term negative behavioral adaptation (e.g., Rajaonah, Tricot, Anceaux, and Millot, 2008) should be studied. The warning application could increase the use of smart phones while driving due to false sense of security, undermining its positive effects. According to the theory of task difficulty homeostasis by Fuller (2005), a driver has a preferred range of driving task difficulty that s/he is prepared to accept and prefers to maintain. Via its warnings, the application could improve the driver's ability to calibrate the perceived situational task demands to her/his capability while multitasking. On the other hand, if the driver's experienced risks of multitasking, informing the estimates of task difficulty, are reduced by the application, the driver could be motivated to increase multitasking behind the wheel. The long-term effects of providing drivers feedback of and support for multitasking behaviors in order to develop their tactical visual off-road sampling and task prioritization skills via a distraction warning system should be carefully studied.

We did not measure directly participants' situation awareness in this experiment but the increase in glance time on road in the most demanding part of the track achieved by the location-based warning could suggest that the warning helped the drivers to recognize the demanding section also with the most demanding in-car task, as they did while conducting the other, easier in-car tasks. Increased attention on road increases the possibilities to detect task-critical events on the road environment compared to a situation where the driver's eyes are off road. However, more glance time on road does not necessarily mean higher situation awareness, and the effects of the warnings on drivers' situation awareness should be more carefully studied in future work. Furthermore, the warning icon was the same for both glance-duration and location-based warnings, and it remains unclear if the drivers were able to correctly associate the location-based warnings to the increased visual demands ahead for each situation.

Due to the unreliable gaze detection, the current study focused mainly on the location-based warnings of the highly demanding parts of the track (intersections, tight bends). This together with the fairly static visual demands of the track on the straights allowed us to test and prove the concept of context-sensitive warnings in a controlled setup, but did not allow us to test the adaptability of the glance duration warning time in more dynamic settings. However, the proactiveness and context-sensitivity of the distraction warnings got significant support from both objective and subjective data.

Whereas in-car user interface design (Lee, Forlizzi, and Hudson, 2005) and testing tools (Kujala and Salvucci, 2015), driver education (Horrey et al., 2009), reactive in-car driver assistance (e.g., lane-keeping assistants) and feedback systems (Donmez et al., 2007), as well as legislative and governmental regulations (NHTSA, 2013b) may help in reducing the negative effects of driver distraction by in-car activities, there is additional demand for fast and cost-effective counter-measures that can be easily deployed by a driver. According to our study, mobile applications aimed to supervise the use of the smart phone while driving and aiding the driver to place more attention on road seems to be a one viable and acceptable option.

Acknowledgments

The authors would like to thank Hilkka Grahn, Niina Holmstedt, Päivi Kinnunen, Jussi Jokinen, Timo Tokkonen, Marko Salmela, Riitta Jokela, Ilkka Kotilainen, and Anton Valojää for their invaluable help in conducting the study. This research was supported by DisGuard, an innovation and research project funded by TEKES (Finnish Funding Agency for Technology and Innovation) and University of Jyväskylä. Jyvässeudun Ajoharjoittelurata Oy provided the test track for the experiment.

References

Baumann, M.R.K., Petzoldt, T., Groenewoud, C., Hogema, J., Krems, J.F., 2008. The effects of cognitive tasks on predicting events in traffic. In Proceedings of the European Conference on HCD for Intelligent Transport Systems 2008 (pp. 3-11).

Bisantz, A.M., Seong, Y., 2001. Assessment of operator trust in and utilization of automated decision-aids under different framing conditions. International Journal of Industrial Ergonomics, 28, 85-97.

Blanco, M., Biever, W.J., Gallagher, J.P., Dingus, T.A., 2006. The impact of secondary task cognitive processing demand on driving performance. Accid. Anal. Prevent. 38, 895-906.

Broström, R., Engström, J., Agnvall, A, Markkula, G., 2006. Towards the next generation intelligent driver information system (IDIS): The Volvo car interaction manager concept. Proceedings of the 13th ITS World Congress.

Donmez, B., Boyle, L., Lee, J.D., 2008. Designing feedback to mitigate distraction. In M.A. Regan, J.D. Lee, and K.L. Young (Eds.), Driver Distraction: Theory, Effects, and Mitigation (pp. 519-532). Boca Raton, FL: CRC Press.

Donmez, B., Boyle, L., Lee, J.D., 2007. Safety implications of providing real-time feedback to distracted drivers. Accid. Anal. Prev., 39, 581-590.

Donmez, B., Boyle, L., Lee, J.D., 2010. Differences in off-road glances: The effects on younger drivers' performance. Journal of Transportation Engineering, 136, 403-409.

Dzindolet, M.T., Peterson, S.A., Pomranky, R.A., Pierce, L.G., Beck, H.P., 2003. The role of trust in automation reliance. International Journal of Human-Computer Studies, 58, 697-718.

Fitch, G.A., Soccolich, S.A., Guo, F., McClafferty, J., Fang, Y., Olson, R.L., Perez, M.A., Hanowski, R.J., Hankey, J.M., Dingus, T.A., 2013. The Impact of Hand-Held And

Hands-Free Cell Phone Use on Driving Performance and Safety-Critical Event Risk (Report No. DOT HS 811 757). Washington, DC: National Highway Traffic Safety Administration.

Flannagan, M., Sivak, M., 1993. Indirect vision systems. In Peacock B. & Karwowski W. (Eds.), Automotive Ergonomics (pp. 205-217). Washington DC: Taylor & Francis.

Fuller, R., 2005. Towards a general theory of driver behaviour. Accident Analysis and Prevention, 37, 461-472.

Ghazizadeh, M., Lee, J.D., Boyle, L.N., 2012. Extending the Technology Acceptance Model to assess automation. Cognition, Technology & Work, 14, 39-49.

Green, P., 2004. Driver Distraction, Telematics Design, and Workload Managers: Safety Issues and Solutions (SAE Paper No. 2004-21-0022).

Grandison, T., Sloman, M., 2000. A survey of trust in internet applications. Communications Surveys & Tutorials, IEEE, 3, 2-16.

Horrey, W.J., Lesch, M.F., 2009. Driver-initiated distractions: Examining strategic adaptation for in-vehicle task initiation. Acc. Anal. and Prev, 41, 115-122.

Horrey, W.J., Lesch, M.F., Garabet, A., 2009. Dissociation between driving performance and drivers' subjective estimates of performance and workload in dual-task conditions. Journal of Safety Res., 40, 7-12.

Jonsson, I.M., Harris, H., Nass, C., 2008. How accurate must an in-car information system be? Consequences of accurate and inaccurate information in cars. In Proceedings of the 26th SIGCHI Conference on Human Factors in Computing Systems (CHI’08) (pp. 1665-1674). ACM.

Jääskeläinen, P., Pöysti L., 2014. Tarkkaamattomuus tieliikenteen turvallisuusongelmana – suomalaisten käsityksiä (Driver distraction and inattention as a safety issue in road traffic – Finnish views). Liikenneturvan selvityksiä 7/2014. Helsinki: Finnish Road Safety Council.

Kaiser, H.F. (1974). An index of factorial simplicity. Psychometrika, 39, 401-415.

Klauer, S.G., Dingus, T.A., Neale, V.L., Sudweeks, J.D., Ramsey, D.J., 2006. The Impact of Driver Inattention On Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data (Report No. DOT HS 810 594). Washington, DC: National Highway Traffic Safety Administration.

Kujala, T., 2013. VisGuard – Distraction Prevention System. Oral presentation at the 3rd International Conference on Driver Distraction and Inattention, Gothenburg, Sweden. Abstract retrieved from http://www.chalmers.se/hosted/ddi2013-en/program/program-papers

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.

Kujala, T., Saariluoma, P., 2011. Measuring distraction at the levels of tactical and strategic control – The limits of capacity-based measures for revealing unsafe time-sharing models. Advances in Human-Computer Interaction, vol. 2011, Article ID 594353, 13 pages.

Kujala, T., Salvucci, D.D., 2015. Modeling visual sampling on in-car displays: The challenge of predicting safety-critical lapses of control. International Journal of Human-Computer Studies, 79. 66-78.

Lee, J., Forlizzi, J., Hudson, S.E., 2005. Studying the effectiveness of MOVE: a contextually optimized in-vehicle navigation system. In Proceedings of the 23th SIGCHI Conference on Human Factors in Computing Systems (CHI’05) (pp. 1665-1674). ACM.

Lee, Y. Lee, J.D., Boyle, L.N., 2007. The interaction of cognitive load and attention-directing cues in driving. Human Factors, 51, 271-280.

Lee, J.D., Moeckli, J., Brown, T.L., Roberts, S.C., Schwarz, C., Yekhshatyan, L., Nadler, E., Liang, Y., Victor, T., Marshall, D., Davis, C., 2013. Distraction Detection and Mitigation Through Driver Feedback. (Report No. DOT HS 811 547A). Washington, DC: National Highway Traffic Safety Administration.

Lee, J.D., See, K.A., 2004. Trust in automation: Designing for appropriate reliance. Human Factors, 46, 50-80.

Lewis, J.R., 1995. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. International Journal of Human-Computer Interaction, 7, 57-78.

Liang, Y., Lee, J.D., Yekhshatyan, L., 2012. How dangerous is looking away from the road? Algorithms predict crash risk from glance patterns in naturalistic driving. Human Factors, 54, 1104-1116.

Mayer, R.C., Davis, J.H., Schoorman, F.D., 1995. An integrative model of organizational trust. Academy of management review, 20, 709-734.

National Highway Traffic Safety Administration, 2013a. Distraction Detection Algorithm Evaluation (Report No. DOT HS 811 548). Washington, DC: National Highway Traffic Safety Administration.

National Highway Traffic Safety Administration, 2013b. Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices (NHTSA-2010-0053). Washington, DC: National Highway Traffic Safety Administration.

Pavlou, P.A., 2003. Consumer acceptance of electronic commerce: integrating trust and risk with the technology acceptance model. International Journal of Electronic Commerce, 7, 101-134.

Rajaonah, B., Tricot, N.J., Anceaux, F., Millot, P., 2008. The role of intervening variables in driver-ACC cooperation. Int. J. Hum.-Comput. Stud., 66, 185-197.

Regan, M.A., Hallett, C., Gordon, C.P., 2011. Driver distraction and driver inattention: Definition, relationship and taxonomy. Acc. Anal. & Prev., 43, 1771-1781.

Schömig, N., Metz, B., 2013. Three levels of situation awareness in driving with secondary tasks. Safety Science, 56, 44-51.

Roberts, S.C., Ghazizadeh, M., Lee, J.D., 2012. Warn me now or inform me later: Drivers' acceptance of real-time and post-drive distraction mitigation systems. International Journal of Human-Computer Studies, 70, 967-979.

Smith, C.A.P., Clegg, B.A., Heggestad, E.D., Hopp-Levine, P.J., 2009. Interruption management: A comparison of auditory and tactile cues for both alerting and orienting. International Journal of Human-Computer Studies, 67, 777-786.

Society of Automotive Engineers, 2000. SAE-J2396 Definitions and Experimental Measures Related to the Specification of Driver Visual Behavior Using Video Based Techniques. Warrendale, PA: SAE.

Tivesten, E., Dozza, M., 2014. Driving context and visual-manual phone tasks influence glance behavior in naturalistic driving. Transportation Research Part F, 26, 258-272.

Transportation Research Board, 2003. Design Speed, Operating Speed, and Posted Speed Practices (NCHRP Report 504).

Underwood, G., Chapman, P., Bowden, K., Crundall, D., 2002. Visual search while driving: skill and awareness during inspection of the scene. Transportation Research Part F, 5, 87-97.

Van Der Laan, J.D., Heino, A., De Waard, D., 1997. A simple procedure for the assessment of acceptance of advanced transport telematics. Transportation Research Part C: Emerging Technologies, 5, 1-10.

Venkatesh, V., Davis, F.D., 2000. A theoretical extension of the technology acceptance model: Four longitudinal field studies. Man. Sci. 46, 186-204.

Wege, C., Victor, T., 2014. Safer attention. Vision Zero International. 28-31.

Weinstock, A., Oron-Gilad, T., Parmet, Y., 2012. The effect of system aesthetics on trust, cooperation, satisfaction and annoyance in an imperfect automated system. Work, 41, 258-265.

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, & S.P. Taylor (Eds.), Vision in vehicles IV (pp. 271-279). Amsterdam: Elsevier.

Wikman, A.S., Nieminen, T., Summala, H., 1998. Driving experience and time-sharing during in-car tasks on roads of different width. Ergonomics, 41, 358-372

You, C.-W., Montes-de-Oca, M., Bao, T.J., Lane, N.D., Lu H., Cardone, G., Torresani, L., Campbell, A.T., 2012. CarSafe: A driver safety app. In Proceedings of UbiComp'12.

Young, K.L., Salmon, P.M., 2012. Examining the relationship between driver distraction and driving errors: A discussion of theory, studies and methods. Safety Science, 50, 165-174.

Young, K.L., Salmon, P.M., Cornelissen, M., 2013. Missing links? The effects of distraction on driver situation awareness. Safety Science, 56. 36-43.

Highlights

- Context-sensitive distraction warnings had a positive effect on drivers' behaviors.

- The warnings significantly increased glance time on road, especially while reading a text message.
- The level of driving experience did not seem to affect the efficiency of the warnings.
- Location-based warnings worked more reliably than glance duration based warnings.
- The context-sensitive warning application was well accepted by the drivers.

Vitae

Tuomo Kujala is a Post-Doctoral Researcher at the Department of Computer Science and Information Technology in the University of Jyväskylä. He obtained his PhD degree in Cognitive Science in 2010 at the University of Jyväskylä, Finland.

Hannu Karvonen is a Research Scientist at VTT Technical Research Centre of Finland Ltd where he works in the Human Factors in Complex Systems team. He is also currently a PhD student in Cognitive Science at the University of Jyväskylä, Finland.

Jakke Mäkelä obtained his PhLic degree in Theoretical Physics in 1998 and PhD degree in Physics in 2010, both at the University of Helsinki, Finland. He is currently a Project Researcher at the Department of Computer Science and Information Technology in the University of Jyväskylä.