JYU DISSERTATIONS 430

Hilkka Grahn

On the Measurement of Visual Distraction Potential of In-Car Activities

UNIVERSITY OF JYVÄSKYLÄ
FACULTY OF INFORMATION TECHNOLOGY

JYU DISSERTATIONS 430

Hilkka Grahn

On the Measurement of Visual Distraction Potential of In-Car Activities

Esitetään Jyväskylän yliopiston informaatioteknologian tiedekunnan suostumuksella
julkisesti tarkastettavaksi
lokakuun 1. päivänä 2021 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Information Technology of the University of Jyväskylä,
on October 1, 2021, at 12 o'clock.

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2021

Editors
Marja-Leena Rantalainen
Faculty of Information Technology, University of Jyväskylä
Päivi Vuorio
Open Science Centre, University of Jyväskylä

Cover photo by Robin Pierre from Unsplash.

ISBN 978-951-39-8841-8 (PDF)
URN:ISBN:978-951-39-8841-8
ISSN 2489-9003

Permanent link to this publication: http://urn.fi/URN:ISBN:978-951-39-8841-8

ABSTRACT

Grahn, Hilkka
On the measurement of visual distraction potential of in-car activities
Jyväskylä: University of Jyväskylä, 2021, 63 p. (+ included articles)
(JYU Dissertations
ISSN 2489-9003; 430)
ISBN 978-951-39-8841-8 (PDF)

People use various applications from Instagram to Netflix while driving. Previous literature recognizes the harmful effects of conducting these secondary in-car tasks while driving. As a general discovery, studies indicate an association between secondary in-car task activities and drivers' visual inattention which is further associated with accidents in traffic. One solution to diminish visual inattention could be to design the user interfaces of the applications to be low-demanding visually and cognitively. However, there is little published data on the exact design factors that could enable such user interface design. There are certain vital issues that complicate studying visual inattention and user interfaces' distraction potential: there is no commonly agreed definition for driver inattention. The lack of an agreed definition leads to difficulties in operationalizing and measuring visual inattention reliably. To be able to define driver inattention, we should first better understand the attentional demands of driving. A more comprehensive understanding of attentional demands of driving could provide instruments that conquer these issues and enable the measurement of visual inattention and examination of the design factors mitigating drivers' visual inattention to enhance traffic safety. Hence, this doctoral dissertation aims to clarify a definition of attentive driving, develop a more reliable method to measure visual inattention, and finally, better understand how user interface design factors affect drivers' visual inattention. This doctoral dissertation makes the following main contributions: a) a suggestion for a working definition of attentive driving, b) an operationalization of visual distraction, c) development of a testing method that assesses tested tasks' visual distraction potential against a baseline of attentive driving and takes drivers' individual glancing behaviors into account, and d) an extension of knowledge concerning the effects of user interface design factors on visual distraction potential. These benefit the traffic research community by helping develop a definition for attentive driving and driver inattention and providing a suggestion of how drivers' visual inattention can be operationalized and measured more reliably. Also, the implications concerning user interface design benefit the automotive industry and designers working within the industry.

Keywords: attentive driving, situation awareness, driver inattention, driver distraction, distraction potential testing, occlusion distance, context-specific design

TIIVISTELMÄ (ABSTRACT IN FINNISH)

Grahn, Hilkka
Ajonaikaisten toissijaisten aktiviteettien visuaalisen tarkkaamattomuuspotentiaalin mittaamisesta
Jyväskylä: Jyväskylän yliopisto, 2021, 63 s. (+ artikkelit)
(JYU Dissertations
ISSN 2489-9003; 430)
ISBN 978-951-39-8841-8 (PDF)

Kuljettajat käyttävät ajaessaan useita sovelluksia Instagramista Netflixiin. Aiempien tutkimusten perusteella tällaiset ajonaikaiset toissijaiset aktiviteetit aiheuttavat tarkkaamattomuutta kuljettajalle, millä puolestaan on tutkimusten mukaan yhteys liikenneonnettomuuksiin. Yksi ratkaisu tarkkaamattomuuden vähentämiseen voisi olla ajonaikaisten toissijaisten sovellusten käyttöliittymien suunnitteleminen niin, että niiden visuaalinen ja kognitiivinen kuormitus olisi mahdollisimman vähäistä. Ongelmana kuitenkin on, että tällaisista suunnitteluratkaisuista on vain vähän tieteellistä tietoa. Kuljettajan tarkkaamattomuuden tutkimista monimutkaistaa myös se, että sille ei ole hyväksyttyä määritelmää tutkijoiden keskuudessa. Määritelmän puute taas johtaa ongelmiin tarkkaamattomuuden operationalisoinnissa ja mittaamisessa luotettavasti. Jotta ylipäätään kuljettajan tarkkaamattomuus olisi mahdollista määritellä hyvin ja luotettavasti, pitäisi ymmärtää paremmin ajamisen vaatimaa tarkkaavuutta. Parempi ymmärrys tarjoaisi kuljettajan tarkkaamattomuuden määritelmän lisäksi instrumentteja tarkkaamattomuuden mittaamiseen liittyvien ongelmien ratkaisemiseen. Tällöin olisi mahdollista myös tutkia luotettavasti käyttöliittymien suunnitteluratkaisuja, joiden avulla voisi olla mahdollista vähentää tarkkaamattomuutta ja tätä kautta parantaa liikenneturvallisuutta. Tämän väitöskirjan tavoite on kirkastaa tarkkaavaisen ajamisen käsitettä, kehittää toissijaisten aktiviteettien tarkkaamattomuuspotentiaalia mittaavaa menetelmää luotettavammaksi sekä lisätä ymmärrystä käyttöliittymien suunnitteluratkaisujen vaikutuksista kuljettajan tarkkaamattomuuteen. Tämän väitöskirjan kontribuutiot ovat: a) ehdotus tarkkaavaisen ajamisen alustavaksi määritelmäksi, b) menetelmä visuaalisen tarkkaamattomuuden operationalisointiin, c) tarkkaavaisen ajamisen avulla määritellyn tarkkaamattomuuspotentiaalia mittaavan ja yksilölliset erot huomioivan testausmenetelmän kehittäminen ja d) lisätiedon tuottaminen käyttöliittymien suunnitteluratkaisujen vaikutuksista kuljettajan visuaaliseen tarkkaamattomuuteen. Väitöskirjan kontribuutiot ovat hyödyllisiä liikenneturvallisuuden tutkijoille määriteltäessä ja mitattaessa kuljettajan tarkkaamattomuutta sekä suunnittelijoille autoteollisuudessa.

Avainsanat: tarkkaavainen ajaminen, kuljettajan tarkkaamattomuus, tarkkaamattomuustestaus, okkluusiomatka, kontekstiin sopiva suunnittelu

Author Hilkka Grahn
Faculty of Information Technology
University of Jyväskylä
Finland

Supervisors Tuomo Kujala
Faculty of Information Technology
University of Jyväskylä
Finland

Pertti Saariluoma
Faculty of Information Technology
University of Jyväskylä
Finland

Reviewers Gustav Markkula
Institute for Transport Studies
University of Leeds
United Kingdom

Gary Burnett
Faculty of Engineering
University of Nottingham
United Kingdom

Opponent Donald L. Fisher
College of Engineering
University of Massachusetts Amherst
United States of America

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest appreciation to my supervisor, Associate Professor Tuomo Kujala for his continuous guidance, encouragement, and patience throughout this dissertation process. Thank you also for initially recommending me for the research assistant job in the cognitive science research group. Your advice on both research and my career in academia has been invaluable. This has been an educational, rewarding, frustrating, exciting, irritating, and fun process. And at least one thing I will remember for sure – "duct tape is needed in every serious data collection process". Thank you.

I am also grateful to my second supervisor, Professor Pertti Saariluoma. Special thanks also to the reviewers of this dissertation, Professor Gustav Markkula and Professor Gary Burnett. Thank you for taking the time to read this dissertation and giving me valuable comments. Special thanks also to my opponent, Professor Donald L. Fisher, for your valuable time and effort. Additionally, I would like to thank my co-authors, Jakke Mäkelä, Annegret Lasch, Johanna Silvennoinen, Aino Leppänen, and Toni Taipalus as well as my research assistants during these years. The numerous participants who took part in our experiments should also be acknowledged; without their volunteer participations, it would have been impossible to conduct these experiments.

All of my colleagues and ex-colleagues, thank you. Special thanks go to Laura Mononen, Markus Salo, Panu Moilanen, Piia Perälä, Tiina Koskelainen, Teija Palonen, Toni Taipalus, Ville Seppänen, and numerous other co-workers. You have taught me much, and you have made me laugh countless times. Special thanks to Panu for patiently teaching me how academia works and how to use särmärit. Thank you for always helping me out and providing me with many interesting job experiences, for instance, in garrisons. And Toni, my co-conspirer, thank you for endlessly supporting me, getting me, making me laugh, and always being there for me. It means the world to me. And thank you for solving the elite culture cues in the crossword puzzles.

Thank you, friends. Our get-togethers and trips with varying themes have helped me forget work and relax and have so much fun. Thank you Mirka, Panu, Tero, Toni, Tupla-A, Ville, Ville, and Wilhelmiina. Special thanks to Ville for listening my whining about, well, pretty much everything during the past years.

My family – dad, Tuomas, Kristina, Perttu, and Jaana – thank you all. Special thanks to my big brother, Tuomas. You have been my role model and inspiration. And thank you, who have already entered into eternity.

Thank you, Gasellit, Mitochondrial Sun, Ramin Djawadi, Rammstein, Soen, and Ulver. Your music has helped me either concentrate, get motivated, cheer up, or to get over the frustration and anger this process has sometimes caused.

Leppävesi 22.8.2021
Hilkka Grahn

TABLE

CONTENTS

LIST OF INCLUDED ARTICLES

I Grahn, H., Kujala, T., Silvennoinen, J., Leppänen, A., & Saariluoma, P. (2020). Expert drivers' prospective thinking-aloud to enhance automated driving technologies – Investigating uncertainty and anticipation in traffic. *Accident Analysis & Prevention, 146*, 105717.

II Grahn, H., & Taipalus, T. (2021). Refining distraction potential testing guidelines by considering differences in glancing behavior. *Transportation Research Part F: Psychology and Behavior, 79*, 23–34.

III 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*, 169–176.

IV 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*, 1–10.

V Grahn, H., & Kujala, T. (2018). Visual distraction effects between in-vehicle tasks with a smartphone and a motorcycle helmet-mounted head-up display. In *Proceedings of the 22nd International Academic Mindtrek Conference*, 153–162.

VI 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*, 102467.

AUTHOR'S CONTRIBUTIONS

Article I

In this article, I was, as a part of the research group, responsible for the formulation of the research goals and aims as well as developing and designing the research methods (that is, hazard prediction test and prospective thinking-aloud method). I prepared the experimental setup with the second author. In Experiment 1 (hazard prediction test), one of the co-authors conducted the experiments while I supervised the experiments. In Experiment 2 (prospective thinking-aloud), I conducted the experiments with one of the co-authors. I was responsible for statistical analyses in Experiment 1. In Experiment 2, the analyses were done in co-operation with two co-authors. Finally, I prepared the initial manuscript (excluding 2.3 Cognitive Mimetics and 4.1.4 Data analysis and partly excluding 4.2 Results and 4.3 Discussion) and subsequently, the co-authors made additions to the manuscript.

Article II

In this article, I was responsible for the line of thought. I performed the conceptualization; that is, formulated the research goals and aims. We used the same data as in Article VI and therefore no new experiments were conducted. I was responsible for the statistical analyses and writing the manuscript. The co-author edited and reviewed the manuscript.

Article III

In this article, I participated in the conceptualization of the research as well as designing the research method. I prepared the experimental setup with the first author. I also conducted the experiments. I was responsible for statistically analyzing the results. I prepared the initial draft of the manuscript under supervision of the first author. Subsequently the co-authors made edits and additions to the manuscript.

Article IV

In this article, I participated in the conceptualization of the research as well as designing the research method. I prepared the experimental setup and conducted the experiments. I was responsible for statistically analyzing the results. I prepared the initial draft of the manuscript under the first author's supervision. Subsequently, the first author made edits and additions to the manuscript. Finally, I presented the research at the *Automotive User Interfaces and Interactive Vehicular Applications* conference.

Article V

In this article, I participated in the conceptualization of the research as well as designing the research method. I prepared the experimental setup with the second author. I supervised the conduction of the experiments, which a research assistant performed. I statistically analyzed the results and prepared the initial manuscript. Subsequently, the co-author made some edits and additions to the manuscript. In addition, I presented the results at the *Academic Mindtrek* conference.

Article VI

In this article, I performed the conceptualization (i.e., formulation of research goals and aims) as well as the design of the research methods under the second author's supervision. I prepared the experimental setup and performed the experiments in both studies. Also, I was responsible for the statistical analyses under the supervision of the second author. Finally, I prepared the initial manuscript and subsequently, the second author made some edits to the manuscript.

1 INTRODUCTION

People use Tinder and Instagram while driving (Ahlström et al., 2019; Kujala & Mäkelä, 2018). People gamble and play augmented reality games behind the wheel (Faccio & McConnell, 2018; Kaviani et al., 2020) and even watch YouTube videos and Netflix (Ahlström et al., 2019; Kaviani et al., 2020; Kujala & Mäkelä, 2018). These studies indicate that smartphones are used for various means while drivers are navigating through busy cities or cruising on highways.

A myriad of studies have examined the detrimental impacts of the use of smartphones, applications, and ubiquitous infotainment systems (later: applications) while driving on traffic safety (e.g., Caird et al., 2014, 2018; Ferdinand & Menachemi, 2014; Guo et al., 2010; Lipovac et al., 2017; Oviedo-Trespalacios et al., 2016; Papantoniou et al., 2017; Simmons et al., 2016, 2017). As a general discovery, these studies indicate an association between application use and drivers' visual distraction. According to one estimation, visual distraction caused by secondary tasks (i.e., tasks not related to driving) accounts for 23 percent of all near-crashes and crashes in the United States (Dingus et al., 2016).

The use of applications while driving might not be such a problem if the human brain were not limited in attending to multiple tasks simultaneously. However, that is not the case; the human brain has limitations in information processing (e.g., Cowan, 2001). Moreover, unfortunately, the user interfaces of applications that drivers use are seldom designed to be low demanding, both visually and cognitively. If these user interfaces were designed well for this safety-critical context (i.e., context-specific design), it could decrease drivers' visual distraction and, hence, enhance traffic safety. In practice, little is still known about the precise design factors of user interfaces that can efficiently diminish drivers' visual distraction.

Visual distraction is a form of visual inattention. However, one complicating factor is the definition of driver inattention or driver distraction – when is the driver actually being inattentive or distracted? For instance, are all glances, that are not directed to the forward road scene, indications of inattention? In addition, both terms – driver inattention and driver distraction – are often used in parallel with each other in the scientific literature. Previous literature has made attempts to define driver inattention and driver distraction. Driver distraction has been defined, for instance, as "any glance that competes with activities necessary for safe driving" (Foley et al., 2013, p. 62) or as "the diversion of attention away from activities critical for safe driving toward a competing activity" (Lee et al., 2009, p. 34). Nonetheless, despite the great number of studies on the matter, there is no agreed definition for driver inattention or driver distraction and, thus, there are numerous ways to operationalize and measure them (e.g., Kircher & Ahlström, 2017). These deficiencies can lead to situations where it is difficult to interpret and compare different research outcomes (e.g., Lee et al., 2009; Pettitt et al., 2005; Regan et al., 2011).

Moreover, there are other issues that hinder the examination of driver inattention and distraction. It has previously been observed that drivers have individual preferences for in-car glance durations (i.e., duration of a glance directed to an in-vehicle application), which seem to be a relatively constant individual tendency (e.g., Broström et al., 2013, 2016; Donmez et al., 2010; Kujala, Mäkelä, et al., 2016; Yang et al., 2021). Several lines of evidence suggest that neglecting these individual tendencies can distort the results of the studies (e.g., Broström et al., 2013, 2016; Lee & Lee, 2017). Again, both, the lack of an agreed definitions for driver inattention and driver distraction, and neglection of individual differences can lead to a situation where interpretation and comparison of the results of inattention and distraction studies are unreliable.

In order to define inattention or distraction, we should better understand the attentional demands of driving. Better understanding the attentional demands of driving could provide us with instruments to measure inattention more reliably and study the effects of secondary in-car tasks on driver inattention in order to enhance traffic safety.

Hence, in this research, the aim is to clarify a definition of attentive driving – and its opposite, inattentive driving – as well as consider how to measure inattention more reliably and better understand the effects of selected in-car task features affecting inattention. Therefore, the following research questions were posited:

1) What is attentive driving?
2) How can driver inattention be measured more reliably and with better validity?
3) What are the effects of selected in-car task features on drivers' visual distraction potential?

The first research question is studied in real traffic with expert drivers using the prospective thinking-aloud method. With the method, it is possible to gain knowledge where expert drivers' attention lie and what the task-relevant events are in a given driving situation. The second and third questions are studied by conducting driving simulator experiments and using eye-tracking technique utilizing a new distraction potential testing method that assesses tested tasks' visual distraction potential against a baseline of attentive driving and takes drivers' individual glancing behaviors into account.

This dissertation consists of four chapters. In Chapter 2, the theoretical foundation of this thesis is reviewed. Next, in Chapter 3, the six articles included in this dissertation are briefly introduced and the contributions regarding this thesis are presented. In Chapter 4, answers to research questions are given reflecting the theoretical foundation. Also, theoretical, methodological, and practical implications of the articles are discussed. Finally, the original articles are included at the end of this doctoral dissertation.

2 THEORETICAL FOUNDATION

This chapter presents the theoretical foundation of this doctoral dissertation. First, in this chapter, attention and situation awareness are examined. Next, inattention and distraction, as well as the measurement of visual distraction potential, are discussed. Finally, the effects of in-car task features on drivers' visual distraction are explored.

2.1 Attention

In order to understand inattention, which is one of this thesis' main concepts, we should first understand attention. If we are able to define what is attentive driving and to what drivers should be paying attention, it could be possible to define when driver inattention occurs (Hancock et al., 2009). Our environment constantly presents more perceptual information than we can efficiently process; therefore, an attentional mechanism is necessary for human beings (Chun et al., 2011). Interest in attention has a long history, from the times of Aristotle (Aristotle, 1957; Hatfield, 1998) to the present day (Wickens, 2021). In the 19^{th} century James (1890) stated that, "Everyone knows what attention is." After 129 years, Hommel et al. (2019) argue, in fact, that even now no one knows what attention is. However, Chun et al. (2011) describe attention as an essential characteristic of all perceptual and cognitive operations that selects, modulates, and sustains focus on information that is most relevant for human behavior, but with a limited capacity. Since attention is incorporated into various human activities from sensory processing to decision-making (Chun et al., 2011), it is a particularly relevant concern in the traffic research (Kircher & Ahlström, 2017).

There are a great number of theories and definitions of attention, such as Broadbent's (1958) Filter model; Treisman's (1960) Filter-attenuation theory; Deutsch and Deutsch's (1963) Late-selection theory; Posner, Snyder and Davidson's (1980) Spotlight theory; and Eriksen and St. James' (1986) Zoom lens model, to name a few. Attention is unwieldy to study (Chun et al., 2011) and therefore, there have been different means of doing so. In some of these renowned pieces of research, attention has been studied with methods of selective listening and a visual search. Later, attention has also been studied with neuroimaging (e.g., Pessoa et al., 2003; Wager et al., 2004). However, in this dissertation, we are interested in attention working in a particular context: driving. We are interested in *to where* and *how much* drivers should direct their attention in order to safely achieve their goals in the driving task. Hence, in this dissertation, we are interested in the targets and contents of attention while driving rather than, for instance, the neural basis of attention. Again, with understanding attention and attentive driving more comprehensively, it could be possible to define inattentive driving.

Regarding this dissertation, Chun et al. (2011) provide a useful taxonomy of attention where they consider attention through a *target* of attention. Chun et al. (2011) argue that attention can be categorized according to the information types that attention operates over; that is, the targets of attention. Therefore, they make a distinction between external and internal attention. External attention selects information coming in through the senses, such as eyes, whereas internal attention selects information, which is represented in the mind, recalled from long-term memory, or maintained in the working memory.

Further, according to Chun et al. (2011), external attention can be subdivided based on the target of attention into sensory modality, spatial locations, time points, features, and objects. Sensory modality refers to vision, hearing, touch, smell, as well as taste and attention then selects and modulates the processing within each of these modalities. In Chun et al.'s (2011) taxonomy, spatial locations refer to spatial attention which prioritizes spatial locations in the environment and is especially, therefore, central to the vision. Often, spatial attention is compared to a metaphor of a spotlight (e.g., Cave & Bichot, 1999; Scholl, 2001). Spatial attention can be both overt [eyes are moved to a relevant location and the focus of attention coincides with the eye movement (e.g., Carrasco, 2011)] or covert [attention is directed to a relevant location without moving the eyes to that location, (e.g., Carrasco, 2011)]. Spatial attention (both overt and covert) can be directed by exogenous (stimulus-driven) and endogenous (goal-directed) cues (Corbetta & Shulman, 2002).

As stated by Chun et al. (2011), time points as a target of attention refers to temporal attention, which share similarities with spatial attention. Temporal attention means that attention is focused on a stimulus that appears in the same location but at different points in time. In other words, this means that attention can be directed to that point in time when a relevant event is supposed to occur in order to optimize behavior (Coull & Nobre, 1998). The amount of environment's objects that can be fully attended to is limited. This means that the information processing rate is limited, and temporal attention therefore selects task-relevant information from the environment to conquer these limitations (Chun et al., 2011). This selecting mechanism of attention applies to other targets of attention, too.

Attention can also be directed at features or objects that can be selected across modality, space, and time (Chun et al., 2011). In Chun et al.'s (2011) taxonomy, features as a target of attention refer to "points in modality-specific dimensions", which are stimuli perceived through modalities, such as color that sticks out, high pitch, or a sudden hot breath of air. Unusual or extreme saliency of the feature has an effect on whether attention is directed to the feature or not. Not just features, but whole objects including all its features can be a target of attention as well (Scholl, 2001).

Internal attention, according to Chun et al. (2011), is targeted at task rules and responses, contents of long-term memory, and contents of working memory. Task rules and responses refer to the choice of a proper response in a selection or decision situation. The contents of long-term memory as a target of attention refers to the determination of which information is encoded into long-term memory and how information is retrieved (Chun & Turk-Browne, 2007). Finally, the contents of working memory as a target of attention refer to maintaining and manipulating information that is no longer externally available. This target can also be referred to as mental representation (e.g., Smith, 1998) and its contents (Saariluoma, 2003). More precisely, the latter refers to the situation-specific information contents of the mental representation.

The taxonomy of attention by Chun et al. (2011) is relevant for the dissertation at hand since it provides a lens through which attention can be seen and it seems to be broad enough to cover attention needed in the complex world of traffic. Hence, here, attention can be understood through the target of attention, both external and internal. External in the sense that information coming through vision is crucial for safe driving since it is estimated that almost 90 percent of the information needed is visual when operating a car (Sivak, 1996). Also, spatial locations – another subcategory of external attention – is relevant since drivers need to prioritize situationally different spatial locations with different weights: for instance, a side view mirror is more important when changing lanes than the speedometer. Another relevant target of external attention is time points; to optimize driving behavior, drivers need to focus their attention on those points in time when a relevant event is expected. For example, when the traffic lights are expected to change. Both spatial and temporal attentions are highly relevant in driving: in terms of safe driving, attention needs to be directed to the relevant locations at the right time. At the same time, internal attention is relevant also: as Chun et al. (2011) argue, internal (or mental) representation (of what is situationally relevant in any traffic situation to attend to), as well as a choice of a proper response in a decision situation, can be targets of attention. The contents of these mental representations are a significant part of safe driving that improve with experience (e.g., Underwood et al., 2002).

2.2 Situation awareness

A concept that is related to attention and is highly relevant from the viewpoint of the taxonomy by Chun et al. (2011) is situation awareness, which could also offer instruments to understand attention in the driving context. Situation awareness (SA) refers to a person's understanding of the state of the environment for succeeding in a task (Endsley, 1995). More accurately, Endsley (1988) has defined situation awareness as follows:

> Situation awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future (Endsley, 1988, p. 792).

According to Endsley (1995), situation awareness has three levels: perception of the elements in the environment (Level 1), comprehension of the current situation (Level 2), and projection of its future status (Level 3). In more detail, achieving Level 1 requires perceiving the status, attributes, and dynamics of the environment's task-relevant elements. In the driving context, achieving Level 1 could mean, for instance, having information on one's own position, other vehicles' positions, and other vehicles' trajectories. Since Level 1 requires environmental sampling, human limitations in visual sampling and attention can lead to errors in Level 1 of situation awareness. This, clearly, can lead to an overall lack of situation awareness.

Furthermore, according to Endsley (1995), achieving Level 2 requires truly understanding the objects and events perceived in Level 1. In the driving context, this could mean needing to understand traffic signs, traffic rules, and how other road users obey those. A novice driver might be able to achieve the same Level 1 situation awareness as an expert, but fails to integrate the data elements in the environment and therefore does not fully understand the ongoing situation. Errors in Level 2 often occur due to the incapability to appropriately comprehend the meaning of perceived data. The misapprehension of the perceived data, or cues, can take place for many reasons, such as when a novice is lacking a mental model [i.e., a structural analogy regarding the real world (Johnson-Laird, 1989)] to comprehend the situation, or when a novice cannot decide which environment's cues are relevant in order to succeed at the task at hand.

Finally, as Endsley (1995) formulated, in order to achieve Level 3, the future states of the environment's elements should be anticipated, allowing for timely and effective decision-making. The anticipation is achieved through the information of the element's status and dynamics as well as understanding the situation. In the driving context, this could, for example, mean the driver's ability to anticipate what is possible to happen if in one's overtaking situation a faster car ahead is approaching slower traffic on the parallel lane. Or overall foreseeing the potential development of future driving situations, such as interruptions in traffic flow, near-accidents or accidents, and acting accordingly. In Level 3, an error can occur if the situation is comprehended accurately, but the future dynamics cannot be anticipated. This error can occur if there is not a highly-developed mental model available. According to Endsley (1995), the ability to acquire situation awareness varies between individuals when they are given the same information input. This is assumed to be due to different information processing mechanisms that are affected by individuals' capabilities, experiences, and training.

Situation awareness has been applied into the driving context also earlier, for instance, by Matthews et al. (2001) and Kaber et al. (2016) who have studied situation awareness in traffic settings. Matthews et al. (2001) presented a model where they integrated Endsley's (1995) situation awareness theory into the goal-oriented model of driver behavior, which includes strategic, tactical, and operational goals of driving. According to Matthews et al. (2001), strategic driving refers to long-term planning of driving, such as, navigating. Tactical driving refers to short-term objectives, such as, a decision when to pass or change lanes. Operational driving refers to translating the tactical decisions into actions to control the vehicle, such as, steering and braking. They conclude that strategic and tactical driving require each level of situation awareness and operational driving requires Levels 1 and 2 in order for the driver to succeed in safely driving. Kaber et al. (2016) concluded that, for successful performance, tactical driving places greater demands on situation awareness than operational and strategic driving. Additionally, the effects of secondary in-car task conduction while driving on situation awareness has been investigated previously too (e.g., Kass et al., 2007; Schömig & Metz, 2013). Kass et al. (2007), for example, noticed that, when engaging in a phone conversation while driving, novice drivers had lower situation awareness than experienced drivers. Schömig and Metz (2013), however, noticed that drivers are able to adjust their interactions with secondary in-car tasks to driving in a situationally aware manner.

It could be argued that situation awareness cannot be achieved without external and internal attention. Achieving Level 1 requires perceiving the environment (Endsley, 1995), which means that external attention selects information coming in through sensory modalities, particularly mainly through vision (Chun et al., 2011). Achieving Level 2 requires comprehending the objects and events (Endsley, 1995), which means that internal attention selects information represented in the mind (Chun et al., 2011) that is based on the driver's previous experiences. Achieving Level 3 requires anticipating the future actions of the environment's objects, which further gives knowledge of how to act upon them (Endsley, 1995). Here, too, internal attention selects information, which is represented in the mind (Chun et al., 2011) and is based on the driver's previous experiences. Internal attention also selects a proper response accordingly (Chun et al., 2011).

In this dissertation, attentive driving is understood through the targets and contents of attention utilizing the theory of situation awareness: the driver is attentive when sufficient Level 3 of situation awareness in the driving task is achieved with appropriate sensory information and mental representations, and is then acted upon.

2.3 Inattention and distraction

There is a substantial amount of scientific literature concerning drivers' inattention and distraction. Despite the number of attempts to define driver inattention or driver distraction, there are no commonly agreed upon definitions (e.g., Foley et al., 2013; Kircher & Ahlström, 2017; Pettitt et al., 2005). Driver inattention has been defined, for instance, by Victor et al. (2009, p. 137) as "improper selection of information, either a lack of selection or the selection of irrelevant information" and Lee et al. (2008, p. 32) defined it as "diminished attention to activities that are critical for safe driving in the absence of a competing activity." Furthermore, Ranney et al. (2000, p. 1) characterize driver distraction simply as "any activity that takes a driver's attention away from the task of driving" and according to Strayer and Fisher (2016, p. 10), driver distraction is "caused by the diversion of attention away from activities critical for safe driving toward activities that are either less critical or unrelated to driving." In addition, Foley et al. (2013, p. 62) have defined visual distraction as "any glance that competes with activities necessary for safe driving." This definition encompasses the expression of "activities necessary for safe driving." It can be argued that Foley et al.'s (2013) definition is incomplete since they do not define the activities necessary for safe driving. In general, this kind of incompleteness creates challenges for operationalizing visual distraction. Operationalization denotes transforming theoretical ideas and intuitions into concrete experimental designs (Saariluoma, 1997).

Additionally, both these terms – inattention and distraction – are used in parallel in the scientific literature. Therefore, it is reasonable to distinguish between inattention and distraction. Regan et al. (2011) have formed a taxonomy of driver inattention where inattention is an umbrella concept and driver distraction (or Driver Diverted Attentions, as labeled in the taxonomy) is one of its subcategories. Driver distraction is further divided into non-driving related (i.e., task-unrelated thoughts, such as daydreaming) and driving-related distraction (i.e., task-related thoughts). Also, for example, Lee et al. (2009) consider distraction as a subset of inattention. The term driver inattention Regan et al. (2011) end up defining as "insufficient, or no attention, to activities critical for safe driving" (p. 1775). However, they also state that what exactly those activities are that are "critical for safe driving" are an unsolved issue. This phrase is, actually, included in many definitions. Driver distraction in the taxonomy by Regan et al. (2011, p. 1776) is defined as "the diversion of attention away from activities critical for safe driving toward a competing activity, which may result in insufficient or no attention to activities critical for safe driving." This "insufficient or no attention to activities critical for safe driving" can be seen as insufficient situation awareness, following the ideas of Endsley (1995): the driver does not perceive the elements in the environment, or at least is not comprehending the current situation, and cannot project its future statuses. However, based on Regan et al.'s (2011) taxonomy, driver distraction is a form of driver inattention and being distracted requires some competing activity. The taxonomy also propounds that a driver can be inattentive while not being distracted, but a driver cannot be distracted without being inattentive. This dissertation adopts this categorization (Regan et al., 2011) of driver distraction as a subcategory of driver inattention.

Overall, these definitions of driver inattention and distraction have been criticized for having a hindsight bias. The hindsight bias refers to defining if the driver was distracted or not *after* knowing the outcome of the driving scenario; meaning that if any kind of accident or performance error occurred or not (Kircher & Ahlström, 2017; Regan et al., 2011). Hence, in order to conquer these hindsight biases in the definitions of driver inattention, we should first define what attentive driving is. Kircher and Ahlström (2017), suggest it is possible to define *minimum attentional requirements* beforehand for different driving situations and maneuvers. These requirements are formulated as rules which must be followed within a particular timeframe. If the requirements are met, the driver is considered attentive. In other words, the dynamically changing demands of different driving situations comprise the minimum requirements for the information that drivers need to sample in order to form and maintain sufficient situation awareness.

Concerning visual sampling, Kircher and Ahlström (2017) suggest that the minimum required attention for every driving situation can be fulfilled with different visual sampling strategies. Later, Ahlström et al. (2021) supplement that the approach of minimum attentional requirements allows drivers to self-regulate their glancing behavior. This means that drivers have a sort of *spare visual capacity* which may be used, for instance, to sample additional information relevant to traffic or to execute secondary in-car tasks (Kircher & Ahlström, 2017). However, those minimum attentional requirements need to be met in order the driver to be classified as attentive. This idea implies that not all off-road glances are equally distractive, as Foley et al. (2013) suggested in their definition of visual distraction, but that the timing of an off-road glance plays a critical role here. Hence, a distracting off-road glance could be interpreted as a calibration failure [i.e., inflated or erroneous estimate of one's own ability or performance (Horrey et al., 2015)] between the momentary visual demands of the driving scenario and the driver's off-road glance length and timing. This interpretation of visual distraction is adopted in this dissertation.

As discussed, there are numerous ways to define driver inattention and driver distraction which also means that there are numerous ways to operationalize and measure them. This makes it difficult to interpret and compare the results of different studies (Regan et al., 2011). Therefore, a well-founded and common definition for visual distraction and its operationalization is needed.

2.4 Measuring visual distraction potential

One way to define driver distraction is to divide it into visual, cognitive, and manual distraction (Foley et al., 2013). However, visual inattention is the most hazardous form of inattention in traffic (e.g., Klauer et al., 2006). Visual inattention is also a form of inattention that can be operationalized and estimated with the eye-tracking technique. Hence, this dissertation is particularly focused on visual inattention. Therefore, only visual inattention, and further, visual distraction are discussed here.

Several lines of evidence suggest that there is an association between drivers' off-road glances and accidents and near-accidents (e.g., Bálint et al., 2020; Dingus et al., 2016). As a result, various authorities have published guidelines on how to assess drivers' visual inattention caused by secondary in-car tasks (e.g., interacting with an application) for industrial testing purposes. For instance, the Alliance of Automobile Manufacturers (AAM, 2006), Japan Automobile Manufacturers (JAMA, 2004), and European Commission (EsOP, 2008) have provided glance durations and glance numbers that should not be exceeded while conducting secondary in-car tasks. Unfortunately, no guidelines were provided on *how* these glance durations and glance numbers should be exactly measured. The first one to do so, was the United States National Highway Traffic Safety Administration (NHTSA, 2013) which published guidelines in 2013 for measuring and assessing how distractive different in-car tasks are.

In these guidelines, distraction potential testing is conducted either using a visual occlusion method or in a driving simulator. In NHTSA's (2013) visual occlusion method, participants complete in-car tasks in a series of 1.5-second glances in a stationary vehicle. To pass the test, the cumulative time of the glances should not exceed 12 seconds. The NHTSA's (2103) occlusion method's capability to measure in-car task's visual distraction potential can be questioned since the method does not involve driving and, hence, is not described here in detail. The NHTSA's (2013) visual occlusion method has been, however, used in previous studies to measure secondary tasks' visual demand, see for instance Burnett et al. (2011). Another testing method presented in the NHTSA guidelines (2013) utilizes a driving simulator. In the method, the testing of distraction potential is conducted in a driving simulator while driving on a straight four-lane road at 50 miles per hour, and following a lead vehicle, and performing secondary in-car tasks. According to the guidelines, testing should be performed with 24 randomly selected participants who are further divided into four groups of six, according to their age (18–24 years, 25–39 years, 40–54 years old, and older than 55 years). Three metrics are used to assess the tested in-car task: total glance time, mean glance duration, and the percentage of over 2-second glances. These metrics mean that (for 21 out of 24 participants):

1) the total glance time should not exceed 12 seconds when performing a task,
2) the mean glance time should be less than or equal to 2 seconds 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.

However, NHTSA's (2013) distraction testing method has received criticism, for instance, for not taking into account the test participants' individual glancing behaviors. This is significant since preceding research indicates that drivers have individual mean in-car glance durations that seem to be relatively constant across tasks (e.g., Broström et al., 2013, 2016; Donmez et al., 2010; Yang et al., 2021). Based on the criticism, Broström et al. (2016) and Ljung Aust et al. (2015) tested how individual glancing behaviors affect the results of the distraction potential testing conducted following the NHTSA (2013) guidelines. They noticed that the results of the distraction potential testing were dependent on the driver sample. This means that the same in-car task with a different driver sample could have had a different outcome in the distraction potential testing. This indicates that if the information on individual glancing behavior is neglected, the results of the distraction potential testing are greatly dependent on the driver sample – not necessarily on how distractive the task at hand is. Hence, the test result can even be false.

In addition, since the driving scenario in the NHTSA (2013) testing method is comprised of a straight four-lane road, another critical observation regarding the method is that it does not account for the visual demands of the driving scenario (e.g., Kujala et al., 2014). That is, the driving scenario in the NHTSA (2013) testing does not correspond sufficiently with the visual demands of real-life driving scenarios (e.g., Kujala et al., 2014; Large et al., 2015), for example, testing a navigation application is rather pointless on a straight road. This is significant since previous research (e.g., Risteska et al., 2021; Tivesten & Dozza, 2014; Tsimhoni & Green, 2001; Wierwille, 1993) has suggested that the visual demands of the driving scenario affect in-car glance durations. For instance, in the study by Large et al. (2015), off-road glances were longer in the NHTSA (2013) scenario than in the more complex scenario. Additionally, the visual demands of driving with different driving simulators, even in a similar scenario, may vary and this can also affect the results of distraction potential testing (e.g., Kujala et al., 2014). These findings suggest that, when conducting distraction potential testing, there is a need for information on how visually demanding certain situations are in the driving scenario. This information would provide a baseline for the accepted glancing behavior in that certain situation, and further, give instruments to assess if the driver is being attentive or not.

In order to respond to the neglects of individual glancing behaviors and visual demands of driving scenarios, Kujala and Mäkelä (2015) introduced a new distraction potential testing method. The new testing method is founded on the occlusion technique, which Senders et al. (1967) initially introduced. Note that this is different from NHTSA's (2013) occlusion technique which does not include driving. In the original technique from Senders et al. (1967), the driver's vision is occluded (i.e., driving blind) and when needed, the driver can see the forward road scene for 500 milliseconds at a time. During the occluded period, the *time* driven without visual information, is measured. Milgram (1987, p. 453) has propounded that, with the occlusion technique, it is possible to "estimate the attentional demand, or information processing workload, imposed on a human monitor/controller of a (complex) system by recording the circumstances and rate at which he/she samples information from the system." Contrary to the original method, in the new testing method by Kujala and Mäkelä (2015), the *distance* driven during the occluded period is measured, not time. This distance is later called the occlusion distance. Occlusion distance stands for the driver's preferred distance in meters that is driven during a period when there is no visual information available. Occlusion distance can also be seen as a measure of the driver's situational spare visual capacity, following Ahlström et al.'s (2021) idea that drivers have a certain amount of time at their disposal to look away from the road scene ahead of them.

In the new distraction potential testing method by Kujala and Mäkelä (2015), the assessment of whether a tested task is too distractive is founded on 97 drivers' occlusion distances (presented in Kujala, Mäkelä, et al., 2016) driven in simulated highways and suburban roads. These occlusion distances were measured and later mapped to the test routes. Each 1x1-meter route point in the map (see Kujala and Mäkelä, 2015) contains information on the median and 85th percentile occlusion distances driven in that particular route point in the original experiment. When the same routes (as in the occlusion distance map) are used later in a distraction potential testing with a new participant sample, this information can be used for categorizing in-car glances as being appropriate or inappropriate glances based on both the distance driven during an in-car glance and the route point where that in-car glance starts.

If the glance is categorized as an appropriate glance, the distance driven during an in-car glance and the visual demands of that route point have been low enough for conducting a secondary in-car task – or a driver has spare visual capacity for conducting an in-car task, as Ahlström et al. (2021) and Kircher and Ahlström (2017) suggest. Low visual demand basically means there are no junctions or sharp road curviness. However, if the in-car glance is categorized as an inappropriate in-car glance, the in-car glance length has exceeded the occlusion distance of the 85th percentile of the original experiment's driver sample (N = 97) on that particular route point. That is, the majority of the original experiment's drivers preferred not to drive in that route point longer without visual information. This means that the in-car glance has been inappropriately long in relation to the visual demands of that given driving situation. These inappropriately long in-car glances are later called *red in-car glances*. In other words, a red in-car glance indicates the driver's visual distraction, and the driver should have been looking at the forward road scene instead of the secondary in-car task on that route point. The idea behind the occlusion distance map is that it can determine the maximum acceptable duration of an in-car glance for each driving situation. This also means that the map provides a baseline for acceptable glancing behavior, which has the same basic idea to first assess attentive driving as in hindsight bias-free minimum attention requirements by Kircher and Ahlström (2017). It should also be noted that the driver can self-pace the acceptable off-road glance duration by speed adjustment, meaning that drivers can regulate the time they drive without visual information while still complying with the acceptable occlusion distance threshold. This is also in line with Kircher and Ahlström's (2017) idea of minimum attentional requirements regarding different, self-regulated sampling strategies to fulfill the minimum requirements for attentive driving.

Other than defining each route point's visual demand, these original occlusion distances of 97 drivers (Kujala, Mäkelä, et al., 2016) are used for validating the new driver sample for the distraction potential testing. With comparing the tested participant sample's occlusion distance distribution to the original occlusion distance distribution of 97 drivers, it is ensured that the new sample matches the original sample and contains drivers with different glancing behaviors – from those drivers who prefer only short occlusion distances to those drivers who prefer longer occlusion distances. A more detailed description of the method can be found in Kujala and Mäkelä (2015).

As argued earlier, there is a need for a more robust visual distraction potential testing method that takes into account both visual demands of the driving scenario and drivers' individual glancing behaviors. The distraction potential testing method of Kujala and Mäkelä (2016) may deliver the need. However, since the method is recent, its reliability and validity should be evaluated. Moreover, this kind of distraction potential testing method together with proper operationalization of driver inattention in relation to a baseline of attentive driving could provide increased comprehension regarding how distracting different in-car task features and interaction methods are for drivers. For the reasons above, the new testing method of Kujala and Mäkelä (2016) is used in the studies included in this dissertation.

2.5 Effects of in-car task features on driver's visual distraction potential

The ample number of technologies used while driving have evoked a number of studies examining the effects of in-car task features on a driver's visual inattention and distraction. Nevertheless, it is not clear which exact user interface design factors have an effect on a driver's visual inattention and how substantial these effects are. However, some general features have been studied which give indications of how distracting they are for drivers. Still, more specific knowledge is needed to understand how these design features affect visual distraction. It should be noted that this is not an extensive review; the studies are selected here for their relevance to this dissertation.

Text entry methods and their effects have been studied earlier. According to the studies (e.g., Crandall & Chaparro, 2012; McKeever et al., 2013; Perlman et al., 2019; Reimer et al., 2014; Tippey et al., 2017; Tsimhoni et al., 2004), text entry with a touch screen keyboard is among the most visually distracting in-car tasks for drivers. Several studies have also indicated that a voice recognition-based text entry (or speech-to-text function) is significantly less distracting than a keyboard text entry (e.g., Beckers et al., 2017; He et al., 2014; He et al., 2015; Tippey et al., 2017; Tsimhoni et al., 2004). However, as Reimer and Mehler (2013) pointed out, it is reasonable to take into account that against common belief, the voice-guided systems usually also include some visual-manual interactions, which may be distractive for drivers.

Another design factor that may diminish a driver's visual distraction is utilizing a read-aloud function as an interaction method. Read-aloud functions read selected text aloud. According to the study of Owens et al. (2011), a read-aloud function is not causing longer off-road glances compared to baseline driving. However, there is not much published research that examines how distracting the read-aloud function (measured with glance duration data) is in the driving context. Conversely to Owens et al. (2011), other studies that are not based on glance durations have concluded that the read-aloud function may not be distraction-free either (Jamson et al., 2004; Lee et al., 2001).

Handwriting is one method to conduct text entries as well. However, there is no extensive literature concerning handwritten text inputs in the automotive context. For example, Burnett et al. (2005) found out that handwriting was a faster text input method than a keyboard (n.b., when writing with a non-preferred left hand). In addition, Kern et al. (2009) studied to where the handwriting surface should be located in the car cockpit. Hence, broader distraction testing of the handwriting method, incorporating in-car glance measurements, seems to be lacking.

In order to reduce drivers' visual distraction, head-up display (HUD) technologies have been in the research focus, too. Head-up displays may have notable potential to reduce visual distraction compared to head-down displays (HDD). For example, Weinberg et al. (2011) noticed that, when using HDD, the number of in-vehicle glances doubled compared to when using HUD. Lagoo et al. (2019) found out that using HUD compared to HDD indicated a 45% improvement in collision avoidance. Topliss et al. (2020) observed that, compared to HUD, HDDs led to a higher percentage of unsafe driving performance. In addition, Villalobos-zúñiga et al. (2016) demonstrated that a combination of a physical keypad and HUD enabled drivers to maintain visual attention on the road up to 64% more compared to a touch screen keyboard. However, HUD may cause negative gaze concentration effects compared to baseline driving without any secondary tasks (Victor et al., 2005).

Intuitively thinking, a bigger screen size enables more efficient task performance and, in general, that has been ensured with studies by Hancock et al. (2015) and Raptis et al. (2013). In the automotive context, Gaffar and Kouchak (2017) studied drivers' reaction times while selecting the target icon on either 7″ or 10″ screen. They did not find differences in reaction times between those two relatively large screens. Unfortunately, they did not measure glance durations. Similarly, previous studies have not extensively dealt with the effects of screen orientation (landscape versus portrait) on visual distraction in detail either, which could be one factor affecting visual distraction. According to one study (Lasch & Kujala, 2012), the screen's orientation had no effect on in-car glance durations.

Another intuitive thought is that bigger button sizes reduce visual distraction. Nevertheless, Feng et al. (2018) concluded that there is no significant difference on driver's visual distraction between medium and large button sizes. A significantly longer total eyes off-road time, however, was found between small button sizes compared to medium or large buttons. Though, it should be noted that these small buttons were smaller (side length of 14 mm) than any recommendation (e.g., Monterey Technologies Inc, 1996) suggest.

According to previous literature, when conducting in-car tasks, page-by-page scrolling with a simple swiping gesture is visually least demanding for drivers compared to button presses or kinetic scrolling (Kujala, 2013; Lasch & Kujala, 2012). In contrary, in Large et al.'s (2013) study where participants were asked to find specific words from a list presented in touch screen, page-by-page scrolling was visually the most distracting, whereas kinetic scrolling was less distracting. However, in Large et al.'s (2013) study, the lists were organized alphabetically which may explain the opposite result as in the previously mentioned studies. When words are in alphabetical order, finding the right word can perhaps be done faster with kinetic scrolling than scrolling page-by-page. If the searched word starts with, for instance, letter T, it is quicker to get there with kinetic scrolling than with page-by-page scrolling. Lasch and Kujala (2012) also concluded that seven items per screen (at least on 4″ screen) is too many items for in-car use. In addition, Kujala and Saariluoma (2011) found out that in-car devices' list-style menu structure led to a smaller variance in in-car glance durations than grid-style menu structure. These findings concerning scrolling methods, number of items, and menu structures could be design factors to diminish drivers' visual distraction.

Well-designed task structures are yet another design factor that could decrease drivers' visual distraction while conducting secondary in-car tasks. Task structure means "how a task breaks down into smaller subtasks" (Salvucci & Kujala, 2016). According to previous observations, people tend to switch tasks at subtask boundaries (e.g., Janssen et al., 2012; Lee et al., 2015; Lee & Lee, 2019; Salvucci & Kujala, 2016), for example typing one word or dialing a phone number in chunks at a time (Janssen et al., 2012) instead of one letter or one number. This attention switching at subtask boundaries reduces cognitive load (e.g., Bailey & Iqbal, 2008; Janssen et al., 2012). If the task structures are designed in a way that enables the use of subtask boundaries, this could be beneficial for the drivers to diminish visual distraction.

The literature on in-car tasks' interaction methods has certainly provided several suggestions on how these general features affect drivers' visual inattention. Typically, these kinds of studies compare the distraction potential of the tested tasks', which provides information on how distractive those tasks might be in relation to each other tested task. Hence, this does not necessarily provide information on the tested task's distraction potential *per se,* compared against a baseline of attentive driving. Further, if the measures used are unreliable, even these relational differences in the task's distraction potential can be false. However, in order to design visually low demanding applications for drivers, more scientific knowledge is needed on how applications' different features and interaction methods affect drivers' visual inattention.

3 RESEARCH CONTRIBUTION

In order to answer the posited research questions, we conducted experiments, scrutinized their results and reported the findings in six articles. This chapter presents those six articles as they form the empirical part of this doctoral dissertation. The original articles are included at the end of this dissertation and therefore only the overviews of the included articles are presented here. The overviews include the aim and the contributions of the articles from the perspective of this dissertation.

3.1 Article I

Grahn, H., Kujala, T., Silvennoinen, J., Leppänen, A., & Saariluoma, P. (2020). Expert drivers' prospective thinking-aloud to enhance automated driving technologies – Investigating uncertainty and anticipation in traffic. *Accident Analysis & Prevention*, *146*, 105717.

As said previously, to understand inattention, we should first understand what attention. Therefore, we set out to more deeply investigate targets of attention in the driving context. To do that, we selected experts in the driving domain – that is, driving instructors – and with the prospective thinking-aloud method examined their anticipations, to what and where these experts attend, and how they act upon in real traffic. The prospective thinking-aloud method was developed precisely for this study.

First, we validated the expertise of the driving instructors with a hazard prediction test (N = 36) in our laboratory to substantiate that they are able to anticipate and predict unfolding hazardous events at a better rate than inexperienced or ordinary (*mixed group* in the article) drivers. After the validation, the subsample of experts (N = 6) drove on public roads while prospectively thinking aloud their anticipations of unfolding traffic events. These anticipations can be seen as sources of uncertainties that are relevant for Level 3 situation awareness (Endsley, 1995).

Regarding research question 1 (*What is attentive driving?*) the major contribution of this article is, that we were able to identify uncertainties that are related to safe driving, situations where these uncertainties arise, and how experts acted to reduce uncertainty. The uncertainties experts (with great experience of driving safely and teaching this to others) raised were triggered by visual cues or visual events which drivers should recognize and act upon in order to maintain safe, comfortable, and economical driving.

3.2 Article II

Grahn, H., & Taipalus, T. (2021). Refining distraction testing guidelines by considering differences in glancing behavior. *Transportation Research Part F: Psychology and Behavior*, *79*, 23–34.

In this article, our objective was to test the robustness of Kujala and Mäkelä's (2015) new distraction testing method, which is used in articles III through VI. Previous studies have shown that the results of the distraction testing conducted following the NHTSA (2013) guidelines can be manipulated by, for instance, altering the participant sample (e.g., Broström et al., 2016; Ljung Aust et al., 2015). In this article, we set out to investigate whether the results of this new distraction potential testing method by Kujala and Mäkelä (2016) can be manipulated in the same way. The NHTSA's testing method is based on measuring static glance metrics while driving on a straight four-lane highway, whereas this new method assesses visual distraction through a baseline of attentive driving founded on the visual demands of that route point where the in-car glance occurs. According to the NHTSA guidelines, testing should be conducted with 24 randomly selected participants who are divided into four groups of six, according to their age. The age groups are 18–24 years, 25–39 years, 40–54 years old, and older than 55.

In the article, we combined data from two experiments (reported in Grahn & Kujala, 2020) that tested the distraction potential of two similar in-car tasks with different participant samples. Both participant samples were initially validated by comparing their occlusion distance distributions against the original occlusion distance distribution of 97 drivers (Kujala, Mäkelä, et al., 2016) to ensure that the samples include participants with different glancing behaviors. These participant samples (N = 23 + N = 23) were then randomly re-organized into ten different participant samples of N = 23 of which occlusion distance distributions were also tested against the original 97-drivers sample. After this, we tested if we were able to produce the same distraction potential test results as in the original study of Grahn and Kujala (2020) with those re-organized participant samples. This was done in order to test whether the new distraction potential testing method by Kujala and Mäkelä (2015) is more robust than testing conducted according to the NHTSA (2013) guidelines – inspired by Broström et al. (2016) and Ljung Aust et al. (2015).

Our results showed that the original experiments' conclusions did not change even though we manipulated the driver samples: with ten different driver samples, the tasks originally labeled distractive remained distractive, regardless of the driver sample. Therefore, we suggest in the article that validating the driver sample with drivers' occlusion distances makes this method more robust and, hence, produces more reliable results than, for example, NHTSA's (2013) method.

According to the NHTSA guidelines, the participant sample should contain people of varying ages. In the article, we suggest that taking into account just the participants' ages is not a sufficient measure for controlling that the driver sample contains drivers with various kinds of glancing behaviors. This suggestion is founded on discovery where, with a hand-picked driver sample containing only drivers with low occlusion distances, we were able to produce a contrary result as in the original experiment: one tested task changed from distractive to non-distractive – even though the hand-picked sample contained drivers from young to old. Hence, we suggest that validating the driver sample with drivers' preferred occlusion distances is a better measure than just leaning on to the assumption that including different age categories in the driver sample is enough to take into account the driver's individual differences in glancing behavior. Overall, this article contributes to research question 2 (*How can driver inattention be measured more reliably and with better validity?*).

3.3 Article III

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*, 169–176.

Our aim in this article was to investigate the distraction potential of an audio-visual navigation system. Three years before this study was published, the National Highway Traffic Safety Administration (NHSTA) released verification guidelines for testing visual-manual in-vehicle devices (NHTSA, 2013). In these guidelines, the testing scenario is a straight highway. However, navigation systems are rarely needed when the route is a straight highway without intersections or turns. Therefore, we argue that the distraction potential of a navigation system cannot be reliably tested with NHTSA's (2013) method. Hence, we utilized a new distraction potential testing method (Kujala & Mäkelä, 2015) with simulated roads that included intersections and turns, in order to adequately test the distraction potential of the route guidance system. To complete the tasks, participants (N = 24) needed to listen to the audio guidance and verify the navigation instructions visually while driving in a driving simulator; no manual input was needed during the tasks. The navigation instructions were presented in two ways: located in the screen's left upper corner or left lower corner, depending on the task.

There are two main findings in this study. First, according to the results, the audio-visual route guidance user interface can be considered having low distraction potential. This finding contributes to research question 3 (*What are the effects of selected in-car task features on drivers' visual distraction potential?*): the tested in-car task's audio-visual interaction method is a feature that has a low effect on the driver's visual distraction.

Second, the used novel distraction potential testing method (Kujala & Mäkelä, 2015) was applied for the first time and the current study provided a baseline for an acceptable in-car task to which more complicated in-car tasks can be compared. The results of the testing implied that the novel distraction testing method is able to produce reliable results, since the results of the testing were similar regardless of the location of the navigation instructions. In addition, the testing was done in a more ecologically valid driving scenario than, for example, the NHTSA (2013) driving scenario. Finally, we introduced the idea that the distraction potential testing methods should include a validation of the driver sample to ensure that it includes all kind of drivers – from short glancing drivers to long glancing drivers. These findings contribute to research question 2 (*How can driver inattention be measured more reliably and with better validity?*).

3.4 Article IV

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*, 1–10.

In this article, we set out to compare the distraction potential of three different text entry methods for in-car tasks: touch screen keyboard, handwriting, and voice recognition. These interaction methods were all features of an automotive-targeted application. To test the distraction potential, we used the same testing method (Kujala & Mäkelä, 2015), as previously. In total, we conducted two experiments in a driving simulator. In the first experiment, we benchmarked the mentioned text entry methods against each other (N = 17). In the second experiment, we investigated more deeply the potential of the handwriting method to be less visually distracting by more thoroughly training the participants to use the method (N = 24). To complete the tasks, participants used the three text entry methods to interact with the automotive-targeted application (e.g., finding an address, making a phone call).

As this article's first contribution, the results suggest that voice recognition is the least visually distracting in-car text input method compared to typing with a touch screen keyboard and a handwriting method. Typing with a touch screen keyboard was the most visually distracting, while there was some promise with the handwriting method to lower the visual distraction compared to keyboard typing. However, we consider that the recognition errors during the handwriting tasks may have affected the results of the visual distraction potential testing. The handwriting method is designed to be used without vision and, therefore, it should have had lower distraction potential. These results contribute to research question 3 (*What are the effects of in-car task features on drivers' visual distraction potential?*): as an interaction method, voice recognition is a feature that seems to be less visually distracting than typing with a touch screen keyboard or typing using the handwriting method.

As this article's second contribution, the results indicate that the used distraction potential testing method (Kujala & Mäkelä, 2015) produces similar results across the tasks: firstly, the distraction potential of voice recognition was similar to the audio-visual route guidance results (low distraction potential) in the previous study (Kujala, Grahn, et al., 2016) as well as in the previous literature (e.g., Beckers et al., 2017; He et al., 2015; Tippey et al., 2017). And secondly, the distraction potential results of the handwriting method were fairly similar in both experiments. Again, this finding contributes to research question 2 (*How can driver inattention be measured more reliably and with better validity?*).

As a minor contribution, we investigated the association between spare visual capacity (measured as occlusion distance) and visual short-term memory capacity (measured with Visual Patterns Test, Della Sala et al., 1999) since Senders et al. (1967) suggested that the time or distance driven with no vision is based on information decay on the forward road scene. The result of this study implies that there is no association between visual short-term memory capacity and the distance a driver is able to safely drive without vision. This finding contributes to research question 1 (*What is attentive driving?*).

3.5 Article V

Grahn, H., & Kujala, T. (2018). Visual distraction effects between in-vehicle tasks with a smartphone and a motorcycle helmet-mounted head-up display. In *Proceedings of the 22nd International Academic Mindtrek Conference,* 153–162.

The aim of this article was to compare the distraction potential between an application that was designed for driving context and regular smartphone applications. Again, we used the same distraction testing method from Kujala and Mäkelä (2015) as previously. In order to complete the tasks, participants (*N* = 24) conducted different search tasks (e.g., searching for a song or contact information) using both applications: an application designed for driving context as well as regular smartphone applications while driving in a driving simulator. The application designed for the driving context utilized helmet-mounted HUD (head-up display) technology together with a steering wheel-mounted controller:

the application view was projected into HUD and the participant could use the application with the controller enabling the use of peripheral vision.

As one of the main contributions of this article, the results suggest that the context-specific design (i.e., a user interface designed specifically for the automotive context) of the HUD application is able to lower the distraction potential compared to regular smartphone applications. This result contributes to research question 3 (*What are the effects of selected in-car task features on drivers' visual distraction potential?*): as an interaction method, HUD technology with a wheel-mounted controller seems to be less visually distracting than interacting with regular smartphone applications. Overall, the context-specific design may have a diminishing effect on a driver's visual inattention.

This article contributes to research question 2 (*How can driver inattention be measured more reliably and with better validity?)* as well. The results of the distraction potential testing are well in line with previous findings in similar tasks with other participant samples (Kujala, Grahn, et al., 2016; Kujala & Grahn, 2017), giving reliability for the used testing method by Kujala and Mäkelä (2015).

3.6 Article VI

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*, 102467.

In this article, we examined more deeply whether context-specific design can have a diminishing effect on visual distraction. Therefore, we set out to study how the interface's interaction methods, in-car task's subtask boundaries, as well as the size of a touch screen affect an in-car tasks' visual demand and visual distraction potential using Kujala and Mäkelä's (2015) testing method. The article consists of two experiments (N = 24 + N = 24) which results are considered both separately and jointly. In the experiments, participants conducted different tasks (e.g., searching a song, reading an email, writing an email) with the automotive-targeted application and with regular smartphone applications while driving in a simulator. The tasks conducted with the automotive-targeted application were executed utilizing speech-to-text and read-aloud functions together with button presses or with simple swiping gestures. The tasks conducted with smartphone applications were executed utilizing a touch screen keyboard or with button presses.

There are several contributions in this article. The overall result, analyzed with multilevel modeling, was that the context-specific design was able to diminish visual distraction: besides the lower distraction potential, tasks conducted with the automotive-targeted application had also lower visual demand than the tasks conducted with regular smartphone applications. According to the multilevel modeling, the automotive-targeted application significantly decreased in-car glance durations compared to regular smartphone applications. We also noted that, perhaps against common belief, bigger screen size had only a minor effect on diminishing the duration of an in-car glance. Based on this, we suggest that the interaction methods of the application are more crucial than merely, for instance, the size of the screen the tasks are conducted with. In addition, we concluded that the orientation of the screen had no effect on drivers' visual distraction.

We also found a plausible impact of subtask boundaries on tested tasks' visual demand and visual distraction. If a task could be divided into smaller subtasks that are determined at the user interface level, together with speech-to-text and read-aloud functions, this may decrease the task's visual demand and drivers' visual distraction.

In this article, we were also able to identify three task groups based on their visual demand: visually high demanding, visually intermediate, and visually low demanding tasks. All of these task groups have their own generalized features, which are presented in Table 1.

TABLE 1: Features of the task groups

Features of visually high demanding tasks	**Features of visually intermediately demanding tasks**	**Features of visually low demanding tasks**
Touch screen typing, self-selected subtask boundaries	Speech-to text function, read-aloud function, subtask boundaries determined from the user interface	Simple swiping gestures

The findings above contribute to research question 3 (*What are the effects of selected in-car task features on drivers' visual distraction potential?)*: as an interaction method, speech-to-text and read-aloud functions seem to diminish the visual demand and distraction potential of the tested tasks compared to regular smartphone applications, well-designed subtask boundaries may decrease drivers' visual distraction, and simple swiping gestures are visually low demanding for drivers. In addition, with multilevel modeling, we were able to indicate that, overall, the context-specific user interface design is capable of diminishing in-car glance durations.

In addition, we suggest a dissociation between visual demand and visual distraction should be made. We justify this suggestion with the observation that some tasks required a high number of in-car glances to be completed – even though the measured visual distraction potential of the task was low. This indicates that the mean number of glances is not alone a sufficient metric for assessing in-car task's visual demand or visual distraction since the visual demands of the driving situation have an impact on glance durations and how distractive the particular in-car glance is. Hence, even if increasing visual demand of the task (measured with total in-car glance durations or number of glances) may increase task's visual distraction potential, visual demand of the task and visual distraction caused by the task are not inevitably congruent. This observation contributes to research question 2 (*How can driver inattention be measured more reliably and with better validity?)*: driver inattention can be measured more reliably by taking into account the visual demands of the driving scenario. In addition, to our best knowledge, this was the first study to analyze tasks' visual demand *while* the variable situational demands of the driving scenario were controlled.

4 DISCUSSION

The aim of this doctoral dissertation was to clarify the definition of attentive driving and to examine how to measure inattention more reliably and with better validity based on this definition. In addition, one aim was to better understand the effects of in-car tasks' features (e.g., text input methods, subtask boundaries) on drivers' visual inattention based on the operationalization presented in this dissertation. Therefore, initially three research questions were posited:

1) What is attentive driving?
2) How can driver inattention be measured more reliably and with better validity?
3) What are the effects of selected in-car task features on drivers' visual inattention?

This chapter presents answers to these research questions. It also considers the theoretical, methodological, and practical contributions drawn from the included articles. The contributions are discussed reflecting the theoretical foundation.

4.1 Theoretical implications

4.1.1 Working definition of attentive driving

According to existing literature, there is no commonly agreed upon definition of driver inattention (e.g., Foley et al., 2013; Kircher & Ahlström, 2017). If we want to define driver inattention and when it occurs, we must first know what attentive driving is and how to define it. Hence, this dissertation's first theoretical implication is a suggested working definition for attentive driving which could further facilitate the traffic safety research community to define driver inattention.

In Article I (Grahn et al., 2020), by studying the mental contents of experts while driving, we were able to identify situation-specific uncertainties that are related to safe (and economic and comfortable) driving. The uncertainties here refer to possible events regarding upcoming driving situations that may or may not unfold. In addition, we studied how experts acted in order to reduce uncertainty. Based on the findings in Article I, it could be considered that attention in the driving context – or attentive driving – could be outlined as follows:

> *Driver recognizes and comprehends those uncertainties in the driving scenario that are relevant for the driving task, and acts accordingly upon to reduce uncertainty into appropriate level, in order to avoid hazardous situations and accidents.*

This suggestion is founded on situation awareness (Endsley, 1995) by means of the taxonomy of external and internal attention (Chun et al., 2011) where attention is categorized according to *targets* of attention.

Based on the theory of situation awareness (Endsley, 1995), attentive driving must achieve three levels: perception of the elements in the environment, comprehension of the current situation, and projection of its future status. Situation awareness, however, cannot be achieved without external and internal attention. External attention is required for directing attention to sensory modality [i.e., mainly vision in driving, see Sivak (1996)] for recognizing and comprehending driving scenario as well as to spatial locations and temporal time points that are essential for driving to prioritize special locations at the right time. Internal attention is required for directing attention to the contents of long-term memory and working memory in order to utilize mental representations or mental models that drivers have gained with driving experience to comprehend and anticipate driving situations, as well as to task rules and responses to choose a proper response in a decision situation. If these are not fulfilled, the driver may be considered inattentive. In Section 2.2, attention in the driving context was outlined as follows: the driver is attentive when sufficient Level 3 of situation awareness in the driving task is achieved with appropriate sensory information and mental representations, and is then acted upon. This is in line with the suggested working definition of attentive driving.

So, combined with these previous theoretical considerations, the suggested working definition of attentive driving can be justified as follows: *Driver recognizes* [external attention, see Chun et al. (2011) and situation awareness Level 1, see Endsley (1995)] *and comprehends* [internal attention, see Chun et al. (2011) and situation awareness Level 2, see Endsley (1995)] *those uncertainties in the driving scenario that are relevant for the driving task* [situation awareness Level 3, see Endsley (1995)], *and acts accordingly upon* [situation awareness Level 3, see Endsley (1995), and internal attention, see Chun et al. (2011)] *to reduce uncertainty into appropriate level in order to avoid hazardous situations and accidents*. A negation of this working definition of attentive driving can serve as a working definition of inattentive driving:

> *Inattentive driving occurs when a driver does not recognize and comprehend those uncertainties in the driving scenario that are relevant for the driving task and, hence, do not act accordingly upon in order to avoid hazardous situations and accidents.*

This theoretical implication answers to research question 1 (*What is attentive driving?)* by providing a working definition for attentive driving.

4.1.2 Dissociation between spare visual capacity and visual short-term memory capacity

In Article IV (Kujala & Grahn, 2017), we noticed that there was no association between spare visual capacity and visual short-term memory capacity. The spare visual capacity was measured with occlusion distance, which refers to a distance that a driver prefers to drive without visual information. The visual short-term memory capacity was measured with the Visual Patterns Test (Della Sala et al., 1999). This observation is contrary to the suggestion that Senders et al. (1967) made, that the time or distance driven without vision is founded on information decay on the forward road scene. This would mean, basically, that the time or distance a driver is able to drive without vision depends on the driver's ability to keep a static picture of the forward road scene in the mind. However, according to our study, the capacity of visual short-term memory was not associated with the drivers' spare visual capacity. Hence, it could be suggested that, during occlusion, drivers are not holding a static image of the road scene in their minds but rather update their dynamic mental representations of the road scene ahead of them.

One can see similarities here with the theory of situation awareness theory (Endsley, 1995): during the unoccluded period, the driver perceives the environment's elements and comprehends the current status (Levels 1 and 2), and during the occluded period, the driver anticipates the future statuses of the environment's elements (Level 3). This theoretical implication answers research question 1 (*What is attentive driving?)* by enhancing the importance of anticipating the future statuses (Level 3) in addition to rather obvious Levels 1 and 2 of situation awareness.

4.1.3 Dissociation between visual demand and visual distraction

In Article VI (Grahn & Kujala, 2020), we suggest, based on empirical data, that a dissociation between secondary in-car task's visual demand and drivers' visual distraction should be made, which is not necessarily clear in the existing literature. Previously, Stevens et al. (2010) defined visual demand as a property of a display or an in-car task – and then distraction is influenced by visual demand and driver's willingness to engage. They also note that visual demand of the in-car task does not necessarily imply driver distraction. However, empirical evidence dissociating visual demand and visual distraction seem to be lacking. The NHTSA (2013) driver distraction guidelines are used to determine if a tested in-car task is *distracting* or not and the mentioned guidelines have been used in various studies examining the *distraction* potential of secondary in-car tasks (e.g., Large et al., 2019; McWilliams et al., 2019; Perlman et al., 2019; Reimer, Mehler, Dobres, et al., 2014). The guidelines are based on static glance metrics, such as, total in-car glance durations and the number of glances. However, in

Article VI, we suggest that NHTSA's (2013) guidelines seem to measure visual *demand* of the tasks, not visual *distraction*.

We justify this suggestion of the dissociation with the observation where some tested in-car tasks required high mean number of in-car glances to be completed even though the measured visual distraction potential of the tasks was low. This means that some tasks required several glances to be completed (i.e., increased visual demand measured with in-car glance durations or number of glances), but drivers were able to time those glances in a way that it did not cause excessive visual distraction for the driver. That is, a task can be visually demanding but not necessarily cause visual distraction. This theoretical implication answers research question 2 *(How can driver inattention be measured more reliably and with better validity?).* With the new distraction potential testing method, we were able to measure both the visual demand of the task and visual distraction caused by the task – not just the visual demand of the task as in, for instance, NHTSA's (2013) method. Additionally, this dissociation is in line with the previously presented working definition of attentive driving. Even though the task is visually demanding, proper timing of the in-car glances can enable drivers to stay attentive, which means that they recognize and comprehend uncertainties in the driving scenario that are relevant for the driving task, and act accordingly upon, in order to avoid hazardous situations and accidents.

Theoretical implications in Section 4.1 benefit the traffic research community by giving ground for developing a definition for attentive driving and driver inattention as well as providing a suggestion for dissociating in-car task's visual demand and visual distraction. These implications also provide a new perspective for the popular idea of Senders et al. (1967) concerning the information decay during occluded driving or glancing off-forward.

4.2 Methodological implications

4.2.1 Operationalization of visual distraction – red in-car glances

The distraction potential testing method used throughout the articles (excluding Article I) utilizes *red in-car glance percentages* to determine if the driver is inattentive, and thereby provides a way to operationalize visual distraction. These red in-car glances are based on how visually demanding the particular road point is where the driver decides to look at the in-car task instead of the forward road scene. The visual demands of a particular route point were estimated with the occlusion distance of 97 drivers in the study by Kujala, Mäkelä, et al. (2016), which provided a baseline for acceptable glancing behavior. This baseline determines, in other words, if the driver is attentive enough, or if the secondary in-car task has caught the driver's attention for too long in a route point where the majority of the 97 drivers chose to see the forward road scene.

According to Kircher and Ahlström (2017), a driver is distracted when the information intake is not meeting the minimum requirements of a given driving situation. This leads to a situation where the driver cannot form a proper internal representation of the driving situation at hand and about to unfold. There are some similarities and associations between the minimum attentional requirements (Kircher & Ahlström, 2017) and situation awareness (Endsley, 1995). *Information intake* is a similar idea to Endsley's (1995) Level 1 (perception of the elements in the environment) in situation awareness theory and *lack of proper internal representation* is similar to Level 2 (comprehension of the current situation) and Level 3 (projection of future statuses), of the same theory. Inspired by these, we conclude that *a red in-car glance can be interpreted as a failure to reach the minimum required attention in a particular driving situation* – which essentially means visual distraction. Hence, that is our suggestion for operationalization of a driver's visual distraction. This conclusion differs from, for instance, NHTSA's (2013) operationalization which is based on static glance metrics, whereas the method by Kujala and Mäkelä (2015) used in Articles II to VI provides a baseline for attentive driving and assesses tested tasks' visual distraction potential against the baseline. However, it should be noted that the used distraction potential testing method cannot necessarily measure Levels 2 and 3 of situation awareness (Endsley, 1995), since it is based solely on glances and their directions. However, if the driver is able to time the in-car glances right, it may be an indication of the appropriateness of the driver's situation awareness in Levels 2 and 3.

Based on the studies conducted for this dissertation, it could be suggested that the red in car-glances used for determining a driver's visual distraction can occur via two mechanisms. Firstly, when looking at the forward road scene, the driver makes a decision if this is a proper time point to engage quickly with the secondary in-car task. When engaging, some feature of the secondary task may catch the attention of the driver for too long; for instance, a natural breakpoint takes longer to occur than expected or a function of the interface is too complicated. This can lead to an in-car glance duration that is too long in relation to the demands of the driving scenario and to what was intended. In this scenario, the mechanism leading to a red in-car glance is, that the driver has accurate situation awareness of the visual demands of the driving situation when the glance begins but some factor in the secondary task catches the driver's attention for too long.

Another possible mechanism of the red in-car glance is, when the driver is looking at the forward road scene, the secondary task causes cognitive load for the driver and disrupts the driver by forming a proper situation awareness of the upcoming visual demands of the driving situation. That is why the driver decides to engage with the secondary in-car task even though the visual demands of the driving scenario are too high. Again, this leads to a too long in-car glance duration in relation to the demands of the driving scenario. In this scenario, the mechanism leading to a red in-car glance is not having an appropriate situation awareness of the upcoming visual demands of the driving situation when engaging with a secondary task due to cognitive distraction. These methodological implications answer research question 2 (*How can driver inattention be measured more reliably and with better validity?)* by providing an operationalization of visual distraction which is founded on a baseline of attentive driving.

4.2.2 Prospective thinking-aloud method

Another methodological implication of this dissertation is a research method called prospective thinking-aloud, which was developed for and utilized in Article I (Grahn et al., 2020)**.** The prospective thinking-aloud method provides an instrument for studying the contents of expert drivers' mental representations concerning traffic situations. With the method, we were able to reveal elements of expert drivers' situation awareness from Level 1 to Level 3 (Endsley, 1995). This methodological implication provides answers to research question 1 (*What is attentive driving?).*

These methodological implications presented in Section 4.2 can be utilized when examining aspects of driver attention and inattention. Besides the operationalization of driver distraction, the idea behind the operationalization presented here can serve as an inspiration to other researchers to develop distraction potential testing methods which determine driver distraction against attentive driving. Moreover, other than studying aspects of situation awareness, the prospective thinking-aloud method can be utilized to develop automated driving technologies to be more "human-like" with the experience and insights of the domain's human experts. In addition, the method can be capitalized on other research fields too, such as aviation.

4.3 Practical implications

4.3.1 Validation of a new distraction potential testing method

Broström et al. (2013, 2016) and Ljung Aust et al. (2015), for instance, have argued that a robust distraction testing method is needed to assess more reliably the distraction potential of secondary in-car tasks. One major practical implication of this dissertation is the validation of the new distraction potential testing method.

The used testing method by Kujala and Mäkelä (2015), to the best of our knowledge, is the first distraction potential testing method that assesses driver distraction against a baseline of attentive driving, or more precisely, against spare visual capacity in attentive driving. This means that, when conducting distraction potential testing with the method, it is possible to determine if the driver should be glancing at the forward road scene instead of the tested in-vehicle device or application on a particular route point. Based on this, it is further possible to assess the distraction potential of the tested task by examining whether the glance durations of the driver are timed right in relation to the variable visual demands of the driving scenario. In other words, the method utilizes a baseline for attentive driving in which the glancing behavior of the distraction potential testing can be reflected against. The driving scenario of this method is also more realistic (containing intersections and curves, for example) than the driving scenario suggested in the NHTSA (2013) guidelines, and drivers can self-pace their off-road glance durations in relation to driving demands with speed adjustment, as in real traffic.

Article III (Kujala, Grahn, et al., 2016) presents the first research that applied this method, and provided a baseline for an acceptable in-car task to which more complicated in-car tasks can be compared. An acceptable in-car task is a task that seems to not cause excessive visual distraction. In later studies (Grahn & Kujala, 2018; Kujala, Grahn, et al., 2016; Kujala & Grahn, 2017), we were able to demonstrate that the distraction potential testing method produced similar results across similar tasks. The similarity of the results is an indication of consistency, which gives the results and the overall method more reliability (Lazar et al., 2010).

The used testing method, to the best of our knowledge, is also the first method that takes into account drivers' individual glancing behaviors or, in other words, takes into account drivers' individual differences. In articles III to VI, we suggest that, in order to take into consideration the drivers' individual differences in glancing behavior, the driver sample should be validated measuring drivers' occlusion distances to ensure that the sample contains drivers who prefer to drive only short distances without visual information to those who prefer to drive long distances without visual information. This idea was further justified in Article II (Grahn & Taipalus, 2021) by suggesting that this kind of procedure improves the robustness of the distraction potential testing. For instance, Ljung Aust et al. (2015) showed that, by manipulating the participant pool, the distraction potential test results following the NHTSA (2013) guidelines had "near stochastic outcomes." In addition, Broström et al. (2016) as well as Lee and Lee (2017) were able to affect the results of the distraction potential testing conducted following the NHTSA (2013) guidelines. In Article II, we demonstrated that, when using the distraction potential testing method by Kujala and Mäkelä (2015), the driver sample does not affect the results of the distraction testing. We suggest that the driver sample validation with occlusion distances is a significant factor in enhancing the robustness of a distraction potential testing method.

In the literature, both neglecting the visual demands of the driving scenario and individual differences in glancing behavior have been raised as major drawbacks in the current distraction potential testing (e.g., Broström et al., 2016; Kujala et al., 2014; Tivesten & Dozza, 2015). It could be argued, based on the studies included in this dissertation, that with similar distraction potential testing method as the method used here (Kujala & Mäkelä, 2015), these recognized drawbacks can be conquered. Hence, Kujala and Mäkelä's (2015) testing method has four benefits compared to, for instance, previously introduced NHTSA's (2013) testing method: 1) it provides a baseline for attentive driving by incorporating the visual demands of the driving scenario which is used to assess the distraction potential of the tested task, 2) the driving scenario is more realistic, containing intersections and curves, 3) drivers can self-pace their off-road glance durations in relation to the driving demands with speed adjustment, as in real traffic, and finally, 4) it takes into account that drivers' have individual glancing behaviors by controlling that there are diverse drivers in the participant sample.

Hence, this practical contribution answers research question 2 (*How can driver inattention be measured more reliably and with better validity?)* by presenting the idea of measuring inattention caused by the secondary in-car tasks assessing tested tasks' visual distraction potential against a baseline of attentive driving and taking drivers' individual glancing behaviors into account.

4.3.2 Context-specific design diminishes visual distraction

Previous research has examined different interaction methods used while driving and conducting a secondary in-car task. Another practical implication of this dissertation is the conclusion that a context-specific user interface design (i.e., a user interface designed specifically for the automotive context) has the potential to diminish drivers' visual distraction. The interaction methods especially seem to have a large effect on drivers' visual distraction. Overall, we were able to produce similar results concerning interaction methods as previous literature: touch screen keyboard is relatively the most distracting interaction method (e.g., Crandall & Chaparro, 2012; McKeever et al., 2013; Reimer, Mehler, & Donmez, 2014), voice-based interaction methods (speech-to-text function and read-aloud function) are less distracting than manual text entry (e.g., Beckers et al., 2017b; He et al., 2014; He et al., 2015; Tsimhoni et al., 2004), simple swiping gestures are visually low distractive, and a head-up display is less distracting than head-down display (e.g., Smith et al., 2016). Additionally, we noticed that the head-up display did not cause gaze concentration. Gaze concentration, or a narrowing of the visual scanning behavior, decreases the ability to detect peripheral and central targets (Wang et al., 2014), which, naturally, affects driver's situation awareness. As a novel discovery, a rather new text entry method called handwriting was found to be less or equally distracting as touch screen keyboard typing. Also, the size of the screen alone had only a minor effect and the orientation of the screen had no effect on visual distraction. These results overall provide more insightful knowledge concerning interaction methods used in user interfaces of different applications. For instance, the knowledge concerning the distraction potential of read-aloud function seems to be especially lacking.

Above all, utilizing multilevel modeling, we were able to conclude that a context-specific design with its multimodal interaction and simplistic design has a diminishing effect on drivers' visual distraction. This means that if an application is designed in the first place to be visually less distracting, bearing in mind the context it is designed for, it indeed has the potential to be visually less distracting. Hence, we suggest that scientific knowledge regarding human–technology interaction should be utilized when designing for a safety-critical context. All these practical contributions concerning interaction methods and context-specific design answer to research question 3 (*What are the effects of selected in-car task features on drivers' visual distraction potential?).*

4.3.3 Carefully designed subtask boundaries benefit drivers

Furthermore, it is possible to diminish the visual distraction potential of in-car tasks by carefully designing the subtask boundaries utilizing natural breakpoints. Basically, this is one part of the context-specific design. Earlier studies have indicated that people have a tendency to switch tasks in natural breakpoints (e.g., Janssen et al., 2012; Lee & Lee, 2019), such as dialing a chunk of phone numbers at once. Based on our results, the possibility to break down an in-car task into smaller subtasks decreased in-car glance durations. This enables drivers to better adjust their glancing behavior in relation to the demands of the driving scenario. This practical contribution answers research question 3 (*What are the effects of selected in-car task features on drivers' visual distraction potential?).*

The practical implications presented in Section 4.3 benefit the research community and other instances (such as New Car Assessment Programs, NCAP, Imberger et al., 2020), focusing on traffic safety research by providing a suggestion of how driver inattention can be measured more reliably. These suggestions could also be utilized when developing driver distraction detection algorithms (e.g., Ahlström et al., 2021). In addition, the implications concerning the design of the secondary tasks benefit the automotive industry and designers working within the industry.

4.4 Summary of the main implications

In order to understand, measure, and evaluate driver inattention more reliably and with better validity, the attentional demands of driving should first be comprehended. This dissertation suggests a working definition for attentive driving: *Driver recognizes and comprehends those uncertainties in the driving scenario that are relevant for the driving task, and acts accordingly upon to reduce uncertainty into appropriate level, in order to avoid hazardous situations and accidents.*

Based on the definition together with the theoretical foundation of the research field, this dissertation also suggests an operationalization for inattentive driving by red in-car glances. This dissertation indicates that when the distraction potential of the tested tasks is assessed against the spare visual capacity in attentive driving, the distraction potential of secondary in-car activities can be estimated more reliably and with better validity. In addition, the results of the distraction potential testing can be utilized when designing user interfaces for safety-critical context to diminish drivers' visual distraction to enhance traffic safety. These results benefit the traffic research community, other instances specialized in traffic safety and the automotive industry, and designers working within the industry.

4.5 Limitations and evaluation of the research

Experimental research has enabled many groundbreaking findings in behavioral science (Lazar et al., 2010). However, experimental research, as any other research method, has limitations to note. Driving simulator experiments in Articles II to VI were conducted in the laboratory settings and this can affect the participants: they may behave in a different way than normally and feel stressed for being observed (Lazar et al., 2010). In general, laboratory experiments are a threat to ecological validity. Typically, ecological validity refers to whether or not the observations made in the laboratory can be generalized to natural behavior in the natural world (Schmuckler, 2001). Then again, the research settings in Articles II–VI would have been hazardous and, hence, unethical to conduct in real traffic, outside the laboratory. It also would not have been feasible to control all the needed variables in real traffic. However, to improve the ecological validity, the driving scenario used with self-paced glance timing and speed is more realistic than, for instance, NHTSA's (2013) scenario. Yet, at the same time, there were no other road users in the driving scenario and this thins down the ecological validity, but other road users could have affected the results of the distraction potential testing by being confounding factors.

Generally, the number of participants in the experiments places limitations on the validity. The participant number varied in our experiments from 17 to 48, and there were only six participants in the study conducted in public roads, (Article I). In Articles II to VI, we used a within-subject design where we compared the performance of the same participants under different conditions, thereby enabling smaller sample sizes than between-subject design (Lazar et al., 2010). This number of participants seemed to be enough since we discovered statistically significant differences between the groups and the effect sizes varied from small to large. In addition, in the real-life study with six participants (Article I), we experienced data saturation in those particular traffic scenarios. Data saturation typically refers to the point at which new information or themes cannot be observed in the data (Guest et al., 2006).

Also, the representativeness of our participant samples should be considered. We used convenience sampling in our experiments, which refers to sampling where researchers select a required number of individuals from participants that are conveniently available (Singleton Jr & Straits, 2005). In our case, this meant that we recruited participants via different mailing lists and potential participants signed up (often university students) for the experiment, and by following previously defined guidelines in participant selection (NHTSA, 2013), we chose individuals to take part in our experiments. With the guidelines, we aimed to improve the representativeness of our participant samples, for example, by also selecting older drivers and not just university students. Hence, the age of the participants varied from 18 years to 79 years. Yet another limitation is the results of the distraction potential testing; the obtained results can be generalized only to the tested tasks.

The distraction potential testing method we used in Articles II to VI is founded on occlusion distances. In Kujala, Mäkelä, et al.'s (2016) study, participants were instructed to drive as accurately as possible without vision and as long as they felt comfortable. Hence, those driven occlusion distances were participants' rough subjective estimates of the visual demands of the driving situations. Therefore, it cannot be stated for sure that the participants were really able to estimate their own abilities to drive without vision for as long as they drove. Here, the distraction potential of the tested tasks were examined comparing in-car glance durations to a baseline of attentive driving, which is founded on those 97 drivers' occlusion distances (Kujala, Mäkelä, et al., 2016) on the same routes. Consequently, the notion of the subjective estimates is significant when evaluating the limitations of this dissertation too. Additionally, another unsolved question to ask is, should those in-car glance durations during the distraction potential testing be compared to a participant's own occlusion distances driven on those same routes rather than the driver population in the study by Kujala, Mäkelä, et al. (2016). In addition, the distraction potential we measured in Articles II to VI was particularly *visual* distraction potential. For more comprehensive testing, cognitive distraction potential should also be evaluated.

The limitations presented in this section should be noted when evaluating the studies included in this dissertation. However, for enhancing the overall validity and reliability of this dissertation, the experimental setups, used apparatus, procedures, and data analyses were described transparently and in detail in the included articles.

4.6 Recommendations for further research

The studies and experiments conducted for this dissertation evoke recommendations for further research. As a future research agenda, examining whether the in-car glance durations should be compared to a participant's own occlusion distances (as mentioned in Section 4.5) would be beneficial. In that case, the participants would provide a baseline for attentive driving for themselves, to which their in-car glance durations could be compared against to. Additionally, the factors affecting individual occlusion distances would be fruitful to investigate. The factors could perhaps be, for example, visual search efficiency, the width of useful field of vision, or some personality trait.

Another recommendation for further research is to utilize the prospective thinking-aloud method outside Finland and broaden the research of uncertainties in traffic to other countries and cultures as well. This could indeed reveal new uncertainties that we were not able to identify in traffic of a relatively small city in a limited set of routes.

In this dissertation, only a limited number of in-car tasks and their visual distraction potentials were tested. Yet another recommendation for further research would be to conduct reliable distraction potential testing, including visual and cognitive distraction, to other kinds of in-car tasks and interaction methods other than the ones tested here. This would provide more comprehensive knowledge concerning how the design of different in-car tasks and interaction methods affect drivers' visual distraction. In addition, instances that are conducting distraction testing, or considering to develop a distraction rating system (NCAP, see Imberger et al., 2020), could utilize the ideas presented in this dissertation concerning distraction potential testing. Overall, when these distraction potential tests are conducted reliably, assessing visual distraction potential against a baseline of attentive driving, this would improve the safety in traffic for all of us.

YHTEENVETO (SUMMARY IN FINNISH)

Aiempien tutkimusten mukaan kuljettajat käyttävät erilaisia sovelluksia deittipalveluista uhkapelipalveluihin autoa ajaessaan. Useat tutkimukset ovat vahvistaneet tällaisten ajonaikaisten toissijaisten aktiviteettien yhteyden kuljettajan visuaaliseen tarkkaamattomuuteen – mikä puolestaan on tutkimusten mukaan yhteydessä liikenneonnettomuuksiin. Yksi ratkaisu tarkkaamattomuuden vähentämiseen voisi olla sovellusten käyttöliittymien suunnitteleminen niin, että ne olisivat kuljettajalle mahdollisimman vähän kuormittavia sekä visuaalisesti että kognitiivisesti. Sovellusten aiheuttaman visuaalisen tarkkaamattomuuden mittaaminen luotettavasti on kuitenkin haastavaa, koska kuljettajan tarkkaamattomuudelle ei ole tutkijoiden keskuudessa hyväksyttyä määritelmää, eikä sen takia myöskään luotettavaa operationalisointia.

Jotta kuljettajan tarkkaamattomuus olisi ylipäätään mahdollista määritellä hyvin ja luotettavasti, pitäisi ymmärtää ajamisen visuaalista vaativuutta paremmin. Parempi ymmärrys ajamisen visuaalisesta vaativuudesta taas voisi tarjota kuljettajan tarkkaamattomuuden määritelmän lisäksi instrumentteja sen mittaamiseen liittyvien ongelmien ratkaisuun. Tällöin olisi mahdollista myös tutkia luotettavasti niitä suunnitteluratkaisuja, joiden avulla olisi mahdollista vähentää kuljettajan tarkkaamattomuutta ja näin parantaa liikenneturvallisuutta.

Tässä väitöskirjassa tutkittiin mitä on tarkkaavainen ajaminen, miten ajonaikaisten toissijaisten tehtävien aiheuttamaa tarkkaamattomuuspotentiaalia voisi mitata luotettavammin ja miten käyttöliittymien suunnitteluratkaisut vaikuttavat kuljettajan visuaaliseen tarkkaamattomuuteen. Tarkkaavaista ajamista tutkittiin eksperttien avulla liikenteessä ja tarkkaamattomuuspotentiaalia ja suunnitteluratkaisujen vaikutusta tutkittiin ajosimulaattorikokeiden avulla.

Väitöskirjalla on useita kontribuutiota. Tässä väitöskirjassa ehdotetaan tarkkaavaisen ajamisen alustavaksi määritelmäksi seuraavaa: *kuljettaja tunnistaa ja ymmärtää ajonäkymässä olevat, ajotehtävälle relevantit epävarmuudet ja toimii tämän perusteella niin, että epävarmuus laskee hyväksyttävälle tasolle, jotta pystyy välttämään riskitilanteita ja onnettomuuksia*. Tämän alustavan määritelmän ja aiemman teoreettisen pohjan avulla tässä väitöskirjassa esitellään myös menetelmä visuaalisen tarkkaamattomuuden operationalisointiin. Väitöskirjassa myös kehitetään tarkkaavaisen ajamisen alustavan määritelmän avulla toissijaisten aktiviteettien tarkkaamattomuuspotentiaalia mittaavaa ja yksilölliset erot huomioivaa testausmenetelmää. Näiden lisäksi väitöskirja tuottaa lisätietoa erilaisten käyttöliittymien suunnitteluratkaisujen vaikutuksista kuljettajan visuaaliseen tarkkaamattomuuteen.

Väitöskirjassa esitellyt tulokset ja kontribuutiot ovat hyödyllisiä liikenneturvallisuuden tutkijoille määriteltäessä ja mitattaessa kuljettajan tarkkaamattomuutta. Esitellyt kontribuutiot ovat hyödyllisiä myös autoteollisuudelle ja siellä työskenteleville suunnittelijoille: tulokset auttavat suunnittelemaan käyttöliittymistä vähemmän tarkkaamattomuutta aiheuttavia, jotta meillä kaikilla olisi turvallisempaa liikenteessä.

REFERENCES

Ahlström, C., Georgoulas, G., & Kircher, K. (2021). Towards a context-dependent multi-buffer driver distraction detection algorithm. *IEEE Transactions on Intelligent Transportation Systems*. https://doi.org/10.1109/TITS.2021.3060168

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

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

Aristotle. (1957). *Parva Naturalia* (W. D. Ross (ed.)). London: Heinemann.

Bailey, B. P., & Iqbal, S. T. (2008). Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Transactions on Computer-Human Interaction*, *14*(4). https://doi.org/10.1145/1314683.1314689

Bálint, A., Flannagan, C. A. C., Leslie, A., Klauer, S., Guo, F., & Dozza, M. (2020). Multitasking additional-to-driving: Prevalence, structure, and associated risk in SHRP2 naturalistic driving data. *Accident Analysis and Prevention*, *137*, 105455. https://doi.org/10.1016/j.aap.2020.105455

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

Broadbent, D. E. (1958). *Perception and Communication*. Pergamon Press.

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

Burnett, G. E., Lomas, S. M., Mason, B., Porter, J. M., & Summerskill, S. J. (2005). Writing and driving: An assessment of handwriting recognition as a means of alphanumeric data entry in a driving context. *Advances in Transportation Studies*, *SPEC.ISSUE*.

Burnett, G., Sharma, N., & Stevens, A. (2011). Modelling and predicting the visual demand of in-vehicle information systems. *2nd International Conference on Driver Distraction and Inattention*.

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

Caird, J. K., Simmons, S. M., Wiley, K., Johnston, K. A., & Horrey, W. J. (2018). Does talking on a cell phone, with a passenger, or dialing affect driving performance? An updated systematic review and meta-analysis of experimental studies. *Human Factors*, *60*(1), 101–133. https://doi.org/10.1177/0018720817748145

Carrasco, M. (2011). Visual attention: The past 25 years. In *Vision Research 51*(13), 1484–1525). https://doi.org/10.1016/j.visres.2011.04.012

Cave, K. R., & Bichot, N. P. (1999). Visuospatial attention: Beyond a spotlight model. In *Psychonomic Bulletin & Review*, *6*(2).

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A Taxonomy of external and internal attention. *Annual Review of Psychology*, *62*, 73–101. https://doi.org/10.1146/annurev.psych.093008.100427

Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current Opinion in Neurobiology*, *17*(2), 177–184. https://doi.org/10.1016/j.conb.2007.03.005

Commission of the European Communities. (2008). Commission recommendation of 26 May 2008 on safe and efficient in-vehicle information and communication systems: Update of the European statement of principles on human-machine interface. *Official Journal of the European Union*.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215. https://doi.org/10.1038/nrn755

Coull, J. T., & Nobre, A. C. (1998). Where and when to pay attention: The neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *Journal of Neuroscience*, *18*(18), 7426–7435. https://doi.org/10.1523/jneurosci.18-18-07426.1998

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–185. https://doi.org/10.1017/S0140525X01003922

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

Della Sala, S., Gray, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern span: A tool for unwelding visuo-spatial memory. *Neuropsychologia*, *37*(10), 1189–1199. https://doi.org/10.1016/S0028-3932(98)00159-6

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, *70*(1), 80–90. https://doi.org/10.1037/h0039515

Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(10). https://doi.org/10.1073/pnas.1513271113

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

Endsley, M. R. (1988). Situation awareness global assessment technique (SAGAT). *IEEE Proceedings of the National Aerospace and Electronics Conference*. https://doi.org/10.1109/NAECON.1988.195097

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors*, *37*(1), 42–64. https://doi.org/10.1518/001872095779049543

Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, *40*(4), 225–240. https://doi.org/10.3758/BF03211502

Faccio, M., & McConnell, J. (2018). Death by Pokémon GO: The economic and human cost of using apps while driving. *National Bureau of Economic Research*, *87*, 815–849. https://doi.org/10.3386/w24308

Feng, F., Liu, Y., & Chen, Y. (2018). Effects of quantity and size of buttons of in-vehicle touch screen on drivers' eye glance behavior. *International Journal of Human-Computer Interaction*, *34*(12), 1105–1118. https://doi.org/10.1080/10447318.2017.1415688

Ferdinand, A. O., & Menachemi, N. (2014). Associations between driving performance and engaging in secondary tasks: A systematic review. In *American Journal of Public Health*, 104(3). https://doi.org/10.2105/AJPH.2013.301750

Foley, J. P., Young, R., Angell, L., & Domeyer, J. E. (2013). *Towards operationalizing driver distraction*. https://doi.org/10.17077/drivingassessment.1467

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

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. *Proceedings of the 22nd International Academic Mindtrek Conference*, 153–162. https://doi.org/10.1145/3275116.3275134

Grahn, H., Kujala, T., Silvennoinen, J., Leppänen, A., & Saariluoma, P. (2020). Expert drivers' prospective thinking-aloud to enhance automated driving technologies – Investigating uncertainty and anticipation in traffic. *Accident Analysis and Prevention*, *146*, 105717. https://doi.org/10.1016/j.aap.2020.105717

Grahn, H., & Taipalus, T. (2021). Refining distraction potential testing guidelines by considering differences in glancing behavior. *Transportation Research Part F: Traffic Psychology and Behaviour*, *79*, 23–34. https://doi.org/10.1016/j.trf.2021.03.009

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough?: An experiment with data saturation and variability. *Field Methods*, *18*(1), 59–82. https://doi.org/10.1177/1525822X05279903

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

Hancock, P. A., Mouloua, M., & Senders, J. W. (2009). On the philosophical foundations of the distracted driver and driving distraction. In M. A. Regan, J. D. Lee, & K. L. Young (Eds.), *Driver Distraction – Theory, Effects, and Mitigation* (pp. 11–30). CRC Press Taylor & Francis Group.

Hancock, P. A., Sawyer, B. D., & Stafford, S. (2015). The effects of display size on performance. *Ergonomics*, *58*(3), 337–354. https://doi.org/10.1080/00140139.2014.973914

Hannah Topliss, B., Harvey, C., & Burnett, G. (2020). How long can a driver look? Exploring time thresholds to evaluate head-up display imagery. *Proceedings of the 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 9–18. https://doi.org/10.1145/3409120.3410663

Hatfield, G. (1998). Attention in early scientific psychology. *Visual Attention*, *1*, 3–25.

He, J., Chaparro, A., Nguyen, B., Burge, R. J., Crandall, J., Chaparro, B., Ni, R., & Cao, S. (2014). Texting while driving: Is speech-based text entry less risky than handheld text entry? *Accident Analysis and Prevention*, *72*, 287–295. https://doi.org/10.1016/j.aap.2014.07.014

He, Jibo, Choi, W., McCarley, J. S., Chaparro, B. S., & Wang, C. (2015). Texting while driving using Google Glass™: Promising but not distraction-free. *Accident Analysis and Prevention*, *81*, 218–229. https://doi.org/10.1016/j.aap.2015.03.033

Hommel, B., Chapman, C. S., Cisek, P., Neyedli, H. F., Song, J. H., & Welsh, T. N. (2019). No one knows what attention is. *Attention, Perception, and Psychophysics*, *81*(7), 2288–2303. https://doi.org/10.3758/s13414-019-01846-w

Horrey, W. J., Lesch, M. F., Mitsopoulos-Rubens, E., & Lee, J. D. (2015). Calibration of skill and judgment in driving: Development of a conceptual framework and the implications for road safety. *Accident Analysis and Prevention*, *76*, 25–33. https://doi.org/10.1016/j.aap.2014.12.017

Imberger, K., Poulter, C., Regan, M., Cunningham, M., & Paine, M. (2020). Considerations for the development of a driver distraction safety rating system for new vehicles. *Journal of Road Safety*, *31*(2). https://doi.org/10.33492/jrs-d-19-00243

James, W. (1890). *The Principles of Psychology*. New York: Henry Holt.

Jamson, A. H., Westerman, S. J., Hockey, G. R. J., & Carsten, O. M. J. (2004). Speech-based e-mail and driver behavior: Effects of an in-vehicle message system interface. *Human Factors*, *46*(4), 625–639. https://doi.org/10.1518/hfes.46.4.625.56814

Janssen, C. P., Brumby, D. P., & Garnett, R. (2012). Natural break points: The influence of priorities and cognitive and motor cues on dual-task interleaving. *Journal of Cognitive Engineering and Decision Making*, *6*(1), 5–29. https://doi.org/10.1177/1555343411432339

Japan Automobile Manufacturers Association. (2004). *Guideline for in-vehicle systems, Version 3.0.*

Johnson-Laird, P. N. (1989). Mental models. In M. I. Posner (Ed.), *Foundations of Cognitive Science* (pp. 469–499). The MIT Press.

Kaber, D. B., Jin, S., Zahabi, M., & Pankok, C. (2016). The effect of driver cognitive abilities and distractions on situation awareness and performance under hazard conditions. *Transportation Research Part F: Traffic Psychology and Behaviour*, *42*, 177–194. https://doi.org/10.1016/j.trf.2016.07.014

Kass, S. J., Cole, K. S., & Stanny, C. J. (2007). Effects of distraction and experience on situation awareness and simulated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *10*(4), 321–329. https://doi.org/10.1016/j.trf.2006.12.002

Kaviani, F., Young, K. L., Robards, B., & Koppel, S. (2020). Nomophobia and self-reported smartphone use while driving: An investigation into whether nomophobia can increase the likelihood of illegal smartphone use while driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *74*, 212–224. https://doi.org/10.1016/j.trf.2020.08.024

Kern, D., Appelmann, T., Schmidt, A., Pararasasegaran, N., Arnsmann, J., & Piepiera, B. (2009). Writing to your car: Handwritten text input while driving. *Conference on Human Factors in Computing Systems - Proceedings*, 4705–4710. https://doi.org/10.1145/1520340.1520724

Kircher, K., & Ahlström, C. (2017). Minimum required attention: A human-centered approach to driver inattention. *Human Factors*, *59*(3), 471–484. https://doi.org/10.1177/0018720816672756

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*. http://hdl.handle.net/10919/55090

Kujala, T. (2013). Browsing the information highway while driving: Three in-vehicle touch screen scrolling methods and driver distraction. *Personal and Ubiquitous Computing*, *17*, 815–823 https://doi.org/10.1007/s00779-012-0517-2

Kujala, T., & Grahn, H. (2017). Visual distraction effects of in-car text entry methods: Comparing keyboard, handwriting and voice recognition. *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 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. *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 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. *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 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. (2018). Naturalistic study on the usage of smartphone applications among Finnish drivers. *Accident Analysis and Prevention*, *115*, 53–61. https://doi.org/10.1016/j.aap.2018.03.011

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

Kujala, T., & Saariluoma, P. (2011). Effects of menu structure and touch screen scrolling style on the variability of glance durations during in-vehicle visual search tasks. *Ergonomics*, *54*(8). https://doi.org/10.1080/00140139.2011.592601

Lagoo, R., Charissis, V., & Harrison, D. K. (2019). Mitigating driver's distraction: Automotive head-up display and gesture recognition system. *IEEE Consumer Electronics Magazine*, *8*(5), 79–85. https://doi.org/10.1109/MCE.2019.2923896

Large, D. R., Burnett, G., Crundall, E., Lawson, G., & De-Kremer, S.-A. (2013). Measuring the distraction of alternative list-scrolling techniques when using touchscreen displays in vehiclese. *Proceedings of the 3rd International Conference on Driver Distraction and Inattention*, 4–6.

Large, D. R., Burnett, G., Crundall, E., Lawson, G., Skrypchuk, L., & Mouzakitis, A. (2019). Evaluating secondary input devices to support an automotive touchscreen HMI: A cross-cultural simulator study conducted in the UK and China. *Applied Ergonomics*, *78*, 184–196. https://doi.org/10.1016/j.apergo.2019.03.005

Large, D. R., van Loon, E., Burnett, G., & Pournami, S. (2015). Applying NHTSA task acceptance criteria to different simulated driving scenarios. *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 117–124. https://doi.org/10.1145/2799250.2799254

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. *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 41–48. https://doi.org/10.1145/2390256.2390262

Lazar, J., Feng, J. H., & Hochheiser, H. (2010). *Research Methods in Human–Computer Research*. Wiley.

Lee, J. D., Caven, B., Haake, S., & Brown, T. L. (2001). Speech-based interaction with in-vehicle computers: The effect of speech-based e-mail on drivers' attention to the roadway. *Human Factors*, *43*(4), 631–640. https://doi.org/10.1518/001872001775870340

Lee, J. D., Young, K. L., & Regan, M. A. (2009). Defining driver distraction. In M. A. Regan, J. D. Lee, & K. L. Young (Eds.), *Driver Distraction – Theory, Effects, and Mitigation* (pp. 31–40). CRC Press Taylor & Francis Group.

Lee, J. Y., Gibson, M., & Lee, J. D. (2015). Secondary task boundaries influence drivers' glance durations. *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 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

Lipovac, K., Đerić, M., Tešić, M., Andrić, Z., & Marić, B. (2017). Mobile phone use while driving-literary review. *Transportation Research Part F: Traffic Psychology and Behaviour*, *47*, 132–142. https://doi.org/10.1016/j.trf.2017.04.015

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*.

Matthews, M. L., Bryant, D. J., Webb, R. D. G., & Harbluk, J. L. (2001). Model for situation awareness and driving: Application to analysis and research for intelligent transportation systems. *Transportation Research Record*, *1779*(1), 26–32. https://doi.org/10.3141/1779-04

McKeever, J. D., Schultheis, M. T., Padmanaban, V., & Blasco, A. (2013). Driver performance while texting: Even a little is too much. *Traffic Injury Prevention*, *14*(2), 132–137. https://doi.org/10.1080/15389588.2012.699695

McWilliams, T., Mehler, B., Seppelt, B., & Reimer, B. (2019). Driving simulator validation for in-vehicle human machine interface assessment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 2104–2108. https://doi.org/10.1177/1071181319631438

Milgram, P. (1987). A spectacle-mounted liquid-crystal tachistoscope. *Behavior Research Methods, Instruments, & Computers*, *19*, 449–456. https://doi.org/10.3758/BF03205613

Monterey Technologies Inc. (1996). *Resource guide for accessible: Design of consumer electronics*.

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

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

Owens, J. M., McLaughlin, S. B., & Sudweeks, J. (2011). Driver performance while text messaging using handheld and in-vehicle systems. *Accident Analysis and Prevention*, *43*(3), 939–947. https://doi.org/10.1016/j.aap.2010.11.019

Papantoniou, P., Papadimitriou, E., & Yannis, G. (2017). Review of driving performance parameters critical for distracted driving research. *Transportation Research Procedia*, *25*, 1796–1805. https://doi.org/10.1016/j.trpro.2017.05.148

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

Pessoa, L., Kastner, S., & Ungerleider, L. G. (2003). Neuroimaging studies of attention: From modulation of sensory processing to top-down control. In *Journal of Neuroscience*, *23*(10), 3990–3998). https://doi.org/10.1523/jneurosci.23-10-03990.2003

Pettitt, M., Burnett, G., & Stevens, A. (2005). Defining driver distraction. *World Congress on Intelligent Transportation Systems*.

Posner, Michael I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160–174. https://doi.org/10.1037/0096-3445.109.2.160

Ranney, T., Mazzae, E., Garrott, R., & Goodman, M. (2000). NHTSA driver distraction research: Past, present, and future. *USDOT, National Highway Traffic Safety Administration*.

Raptis, D., Tselios, N., Kjeldskov, J., & Skov, M. B. (2013). Does size matter? Investigating the impact of mobile phone screen size on users' perceived, effectiveness and efficiency. *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '13*, 127–136. https://doi.org/10.1145/2493190.2493204

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

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

Reimer, B., Mehler, B., & Donmez, B. (2014). A study of young adults examining phone dialing while driving using a touchscreen vs. a button style flip-phone. *Transportation Research Part F: Traffic Psychology and Behaviour*, *23*, 57–68. https://doi.org/10.1016/j.trf.2013.12.017

Risteska, M., Kanaan, D., Donmez, B., & Winnie Chen, H. Y. (2021). The effect of driving demands on distraction engagement and glance behaviors: Results from naturalistic data. *Safety Science*, *136*, 105123. https://doi.org/10.1016/j.ssci.2020.105123

Saariluoma, P. (1997). *Foundational Analysis – Presuppositions in Experimental Psychology*. Routledge.

Saariluoma, P. (2003). Apperception, content-based psychology and design. In U. Lindemann (Ed.), *Human Behaviour in Design* (pp. 72–78). Springer.

Salvucci, D. D., & Kujala, T. (2016). Balancing structural and temporal constraints in multitasking contexts. *Proceedings of the 38th Annual Conference of Cognitive Scince Society*.

Schmuckler, M. A. (2001). What is ecological validity? A dimensional analysis. *Infancy*, *2*(4), 419–436. https://doi.org/10.1207/S15327078IN0204_02

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, *80*(1–2), 1–46. https://doi.org/10.1016/S0010-0277(00)00152-9

Schömig, N., & Metz, B. (2013). Three levels of situation awareness in driving with secondary tasks. *Safety Science*, *56*, 44–51. https://doi.org/10.1016/j.ssci.2012.05.029

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*.

Simmons, S. M., Caird, J. K., & Steel, P. (2017). A meta-analysis of in-vehicle and nomadic voice-recognition system interaction and driving performance. *Accident Analysis and Prevention*, *106*, 31–43. https://doi.org/10.1016/j.aap.2017.05.013

Simmons, S. M., Hicks, A., & Caird, J. K. (2016). Safety-critical event risk associated with cell phone tasks as measured in naturalistic driving studies: A systematic review and meta-analysis. *Accident Analysis and Prevention*, *87*, 161–169. https://doi.org/10.1016/j.aap.2015.11.015

Singleton Jr, R. A., & Straits, B. C. (2005). *Approaches to Social Research*. Oxford University Press.

Sivak, M. (1996). The information that drivers use: Is it indeed 90% visual? *Perception*, *25*(9), 1081–1089. https://doi.org/10.1068/p251081

Smith, E. R. (1998). Mental representation and memory. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The Handbook of Social Psychology* (pp. 391–445). Oxford University Press.

Smith, M., Gabbard, J. L., & Conley, C. (2016). Head-up vs. head-down displays: Examining traditional methods of display assessment while driving. *Proceedings of the 8th International Conference on Automotive User Interfaces* *and Interactive Vehicular Applications*, 185–192. https://doi.org/10.1145/3003715.3005419

Stevens, A., Burnett, G., & Horberry, T. (2010). A reference level for assessing the acceptable visual demand of in-vehicle information systems. *Behaviour and Information Technology*, *29*(5), 527–540. https://doi.org/10.1080/01449291003624212

Strayer, D. L., & Fisher, D. L. (2016). SPIDER: A aramework for understanding driver distraction. *Human Factors*, *58*(1), 5–12. https://doi.org/10.1177/0018720815619074

Tippey, K. G., Sivaraj, E., & Ferris, T. K. (2017). Driving while interacting with Google Glass: Investigating the combined effect of head-up display and hands-free input on driving safety and multitask performance. *Human Factors*, *59*(4), 671–688. https://doi.org/10.1177/0018720817691406

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

Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, *12*(4), 242–248. https://doi.org/10.1080/17470216008416732

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

Tsimhoni, O., Smith, D., & Green, P. (2004). Address entry while driving: Speech recognition versus a touch-screen keyboard. *Human Factors*, *46*(4), 600–610. https://doi.org/10.1518/hfes.46.4.600.56813

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: Traffic Psychology and Behaviour*, *5*(2), 87–97. https://doi.org/10.1016/S1369-8478(02)00008-6

Victor, T. W., Engström, J., & Harbluk, J. L. (2009). Distraction assessment methods based on visual behavior and event detection. In M. A. Regan, J. D. Lee, & K. L. Young (Eds.), *Driver Distraction – Theory, Effects, and Mitigation* (pp. 135–165). CRC Press Taylor & Francis Group.

Victor, T. W., Harbluk, J. L., & Engström, J. A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F: Traffic Psychology and Behaviour*, *8*(2 SPEC. ISS.), 167–190. https://doi.org/10.1016/j.trf.2005.04.014

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

Wager, T. D., Jonides, J., & Reading, S. (2004). Neuroimaging studies of shifting attention: A meta-analysis. *NeuroImage*, *22*(4), 1679–1693. https://doi.org/10.1016/j.neuroimage.2004.03.052

Wang, Y., Reimer, B., Dobres, J., & Mehler, B. (2014). The sensitivity of different methodologies for characterizing drivers' gaze concentration under increased cognitive demand. *Transportation Research Part F: Traffic Psychology and Behaviour*, *26*(PA), 227–237. https://doi.org/10.1016/j.trf.2014.08.003

Weinberg, G., Harsham, B., & Medenica, Z. (2011). Evaluating the usability of a head-up display for selection from choice lists in cars. *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2011*, 39–46. https://doi.org/10.1145/2381416.2381423

Wickens, C. (2021). Attention: Theory, principles, models and applications. *International Journal of Human-Computer Interaction*, *37*(5), 403–417. https://doi.org/10.1080/10447318.2021.1874741

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.*

Yang, S., Kuo, J., & Lenné, M. G. (2021). Individual differences in glance patterns under distraction in level 2 automated driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. https://doi.org/https://doi.org/10.1177/1071181320641477

ORIGINAL PAPERS

I

EXPERT DRIVERS' PROSPECTIVE THINKING-ALOUD TO ENHANCE AUTOMATED DRIVING TECHNOLOGIES – INVESTIGATING UNCERTAINTY AND ANTICIPATION IN TRAFFIC

by

Hilkka Grahn, Tuomo Kujala, Johanna Silvennoinen, Aino Leppänen & Pertti Saariluoma, 2020

Accident Analysis & Prevention, 146, 105717

https://doi.org/10.1016/j.aap.2020.105717

Accident Analysis and Prevention 146 (2020) 105717

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Expert Drivers' Prospective Thinking-Aloud to Enhance Automated Driving Technologies – Investigating Uncertainty and Anticipation in Traffic

Hilkka Grahn*, Tuomo Kujala, Johanna Silvennoinen, Aino Leppänen, Pertti Saariluoma

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

ARTICLE INFO

Keywords:
Automated driving
expert driver
prospective thinking-aloud
traffic safety
uncertainty
anticipation

ABSTRACT

Current automated driving technology cannot cope in numerous conditions that are basic daily driving situations for human drivers. Previous studies show that profound understanding of human drivers' capability to interpret and anticipate traffic situations is required in order to provide similar capacities for automated driving technologies. There is currently not enough a priori understanding of these anticipatory capacities for safe driving applicable to any given driving situation. To enable the development of safer, more economical, and more comfortable automated driving experience, expert drivers' anticipations and related uncertainties were studied on public roads. First, driving instructors' expertise in anticipating traffic situations was validated with a hazard prediction test. Then, selected driving instructors drove in real traffic while thinking aloud anticipations of unfolding events. The results indicate sources of uncertainty and related adaptive and social behaviors in specific traffic situations and environments. In addition, the applicability of these anticipatory capabilities to current automated driving technology is discussed. The presented method and results can be utilized to enhance automated driving technologies by indicating their potential limitations and may enable improved situation awareness for automated vehicles. Furthermore, the produced data can be utilized for recognizing such upcoming situations, in which the human should take over the vehicle, to enable timely take-over requests.

1. INTRODUCTION

Automated driving solutions (i.e., autopilot technologies) are becoming increasingly common in commercial vehicles. The aim of automated driving technology is to substantially decrease accidents and increase driving comfort (Hubmann et al., 2018). Automated driving technologies are sometimes claimed to be safer than human drivers (e.g., McGoogan, 2016; Associated Press, 2018; Teoh & Kidd, 2017), and in many respects they may be superior to a human driver. They are able, for instance, to monitor surrounding objects continuously – automated driving technologies do not get tired or bored during monotonous driving as human drivers tend to do (Horne and Reyner, 1995; Schmidt et al., 2009; Thiffault and Bergeron, 2003; Ting et al., 2008).

One major manufacturer of automated driving technology is Tesla, Inc. and it has been estimated that Tesla's autopilot has driven over 2.2 billion miles on public roads by January 2020 (Friedman, 2020). Nonetheless, these successfully driven kilometers in limited driving scenarios may not be a sufficient indicator of the safety or superiority of these systems over human drivers – automated driving technologies still have some major weaknesses compared to human drivers. Up to now, little attention has been paid to these weaknesses and how human drivers manage in similar situations.

The current traffic system is a social environment where other road users' behavior determines how drivers interact with each other (Zaidel, 1992). Driving is not only a mechanical performance, it is also a "complex social activity" (Brown, 2017). Hence, the interaction between automated driving technology and other road users is gaining attention in the literature (e.g., Brown & Laurier, 2017; Rasouli & Tsotsos, 2019). Schwarting et al. (2018) have stated that interaction between automated driving technologies and human road users is "an unsolved problem". Previous research has identified these problems that automated driving technologies might come across in traffic while interacting with humans – such as lack of negotiation with human drivers (Chater et al., 2018), social issues regarding lane changes and merging (Brown and Laurier, 2017), as well as lack of interaction and communication with pedestrians (Mahadevan et al., 2018). All these studies describing interaction problems between humans and automated driving technologies concluded that these technologies need more "human-like" features to overcome the found social issues.

What could these "human-like" features be? What could explain the

* Corresponding author.
E-mail addresses: hilkka.grahn@jyu.fi (H. Grahn), tuomo.kujala@jyu.fi (T. Kujala), johanna.silvennoinen@jyu.fi (J. Silvennoinen), aino.e.leppanen@student.jyu.fi (A. Leppänen), pertti.saariluoma@jyu.fi (P. Saariluoma).

https://doi.org/10.1016/j.aap.2020.105717
Received 3 April 2020; Received in revised form 3 August 2020; Accepted 3 August 2020

insufficiencies of Tesla's autopilot and other self-driving cars compared to human drivers? Stahl, Donmez, and Jamieson (2013) suggested that we should better understand human drivers' capability to interpret and predict traffic situations to facilitate drivers' competence. Correspondingly, we suggest that one of these "human-like" features that state-of-the-art automated driving technologies lack is the skill of anticipation of traffic events, and more specifically, the skill of recognizing uncertainties concerning the unfolding driving situation and adapting to these accordingly. With this knowledge, automated driving technologies could be trained to perform in a similar way than humans do – or even better.

The aim of this paper is to investigate human experts' uncertainties that rise in anticipatory driving and their related adaptive behaviors. This knowledge is important in order to improve today's automated driving technologies to be safer, more economical, and more comfortable. The research questions are: 1) What are the context-dependent uncertainties that arise in anticipatory driving of expert drivers (here: driving instructors)? and 2) How expert drivers adapt their behavior in the identified driving situations in order to resolve the uncertainties?

First, commercial hazard perception test video clips were analyzed for identifying the situations which may be efficiently anticipated by human drivers but which could cause problems to current automated driving technologies for numerous reasons, such as poor visibility, objects that are partially occluded, unexpected trajectories, or lack of understanding the world. The selected video clips were transformed into hazard prediction clips by ending them with an occlusion just after the hazardous situation started to unfold. These hazard prediction clips were used to test if there are differences between inexperienced, a mixed group of drivers, and driving instructors in anticipating unfolding hazards in traffic. With this experiment, the selected driving instructors' expertise in hazard prediction ability was validated. After the expert sample validation, six of the experts drove a predefined route on public roads while thinking aloud prospectively what driving-task relevant they are anticipating to happen. The research process is illustrated in Fig. 1. Based on the content analysis of the data, uncertainties, as well as related adaptive and social behaviors in specific traffic situations and environments, were identified. To our best knowledge, this paper is the first to investigate human expert drivers' anticipations and uncertainties on public roads with the prospective thinking-aloud method.

2. RELATED LITERATURE

2.1. Automated driving taxonomy and situation awareness

In order to provide understanding of current automated driving technologies and their abilities, the Society of Automotive Engineers (2019) has presented a taxonomy regarding levels of automated driving. Some of today's automated driving technologies may be classified at the third level: automated driving technology can drive the vehicle under limited conditions, and when the system requires, the driver must take over the automated driving technology (SAE, 2019). However, in order to succeed in safe driving already at these levels, the driving task requires situation awareness (Matthews et al., 2001; Ward, 2000).

According to Endsley (1995), situation awareness (SA) refers to understanding the environment's state for succeeding in a task. SA has three levels: perception of the elements in the environment (Level 1), comprehension of the current situation (Level 2), and projection of its future status (Level 3). All levels of driving task (operational, tactical, and strategic) require each level of situation awareness (Matthews et al., 2001).

It can be argued that today's automated driving technology may reach the Level 1 of situation awareness: they recognize environment's elements such as other vehicles, road curviness, and obstacles. But do automated driving technologies reach the Level 2 of situation awareness, comprehension of the current situation? According to Lake, Ullman, Tenenbaum, and Gershman (2017), automated driving technology algorithms can only recognize objects but cannot understand scenes, that is, *comprehend the current situation*.

Situation awareness' Level 3 requires anticipating the future status of the task environment. In the automotive context this means, for instance, predicting other road users' behavior. Predicting other road users' trajectories with different machine learning techniques is, indeed, a growing research area (e.g., lane changes: Chae et al., 2017; Dong et al., 2019; Wissing et al., 2018). According to academic research, state-of-the-art automated driving algorithms may be able to predict trajectories of recognized moving objects when interacting with these objects, selecting optimal paths and speeds accordingly, for instance, in complex intersection scenarios (Hubmann et al., 2018). Meghjani et al. (2019) have developed decision-making algorithms that are able to utilize contextual information (e.g., map data of intersections and lanes ahead) in inferring intentions of the cars in front of the ego vehicle for optimizing lane changes and route planning under uncertainty. However, these fairly low-level and relatively short-term prediction abilities are not yet sufficient when compared to human expert drivers. Lake et al. (2017) point out that – compared to humans – automated driving technologies lack intuitive psychology to be able to anticipate other road users' behavior and intentions. Furthermore, they are lacking in intuitive physics in order to reason about the stability and trajectories of objects that may be occluded momentarily by other objects in the environment. That said, it could be argued that today's autopilots have severe deficiencies at Levels 2 and 3 of situation awareness, which are crucial for safe and comfortable driving (e.g., Baumann & Krems, 2007; Stahl et al., 2013).

2.2. Problems current automated driving technologies encounter on public roads

The literature review on automated technology problems is focused on publications between 2015 and 2020 as the technology is developing rapidly. There are numerous of YouTube videos available where one can see situations in which the driver needs to overtake the automated driving technology (e.g., https://tinyurl.com/yywtj4oo and https://tinyurl.com/y3kae45d). In these videos, Tesla autopilot owners have recorded their drives on public roads while enabling the autopilot. Based on the real-life footage, human intervene is needed, for instance, in situations where lane markings are not clear, when road is too narrow, when ramp is too curvy, or when there are unusual objects on the road.

The lack of scene understanding and future status anticipation may be some of the reasons that have led automated driving technologies to encounter these problems on public roads. One additional component of scene understanding could be the understanding of the social side of traffic. Brown and Laurier (2017) analyzed YouTube video clips of self-driving cars recorded by drivers and documented social challenges that

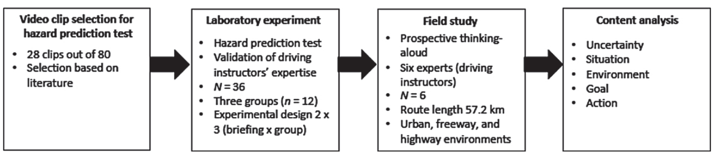

Fig. 1. Research process.

automated driving technologies confront in real traffic. They noticed, for instance, how automated driving technology's lane-changing behavior can be interpreted as rude, how automated driving technology maintaining speed and following traffic lines in merging cause a hazardous situation, and when automated driving technology is not "creeping" in the four-way stop intersection it gets "cut-up" and causes sudden braking. All these actions risk safe, economical, and comfortable traffic flow.

Endsley (2017) studied (based on her own experiences) Tesla's autopilot for a six-month period in 2017. During the period, Tesla's autopilot had problems with sharp turns, curves, merging lanes, and intersections without lane markings. Similarly, Dikmen and Burns (2016) found out in their survey that Tesla's autopilot's lane detection failures caused problems to drivers: the autopilot tried to take an exit ramp and cross lanes for no reason, for example. Endsley (2017) also noticed that while parking, avoiding an obstacle on Tesla's way also led the autopilot into a strange turning angle in a tight place.

According to Dikmen and Burns (2016), one of the current automated driving technologies' major problems is sudden changes in speed. Sudden braking and uncomfortable accelerations and decelerations were brought up in the survey they conducted – these were problems especially in heavy traffic conditions. Naturally, even if not safety-critical, these sudden speed changes are diminishing driving comfort.

Lv et al. (2018) studied automated driving technology manufacturers' reports that summarize incidents when either the technology itself disengaged the autopilot, or the autopilot was disabled by the driver. The latter is called "active disengagement" and means that the automated driving technology does not detect any problem, but the driver notices some unfolding event that makes the driver to take over the control of the car. Active disengagements happened, for instance, when there were too many vehicles and other road users in an intersection; when the automated driving technology did not slow down when a vehicle in front stopped; when the automated driving technology did not recognize a vehicle pulling out from a parking lot; when there was an emergency vehicle on the road or an accident; when other drivers' behavior was unexpected or reckless; and when extra space for a cyclist was needed. All these examples can be considered as situations that diminish safe driving and causes uncertainty of the automated driving technologies' behavior for the driver or passengers.

The reasons causing drivers to disable the automated driving technology have also been studied in a driving simulator. van Huysduynen et al., (2018) noticed in their driving simulator study that the automated driving technology was disabled, for instance, in situations where the technology was perceived as conservative. This means that it reduced speed before passing, and after passing it constantly tried to return to the right lane even if there were slower traffic ahead. Therefore, one recognized reason to disable the automated driving technology was to maintain the traffic flow when changing lanes. Another reason to disable the automated driving technology was due to unpredictability of other road users – drivers did not trust that the technology would cope in those situations. Again, maintaining the traffic flow and coping in uncertain situations are linked to safer, more economical, and more comfortable driving.

In addition, an increasing number of studies have investigated the interaction between automated driving technologies and pedestrians in urban environments. For example, Mahadevan et al. (2018) focused on communication and interaction between automated driving technologies and humans. They emphasized the importance of the communication that the vehicle is aware of pedestrians. This kind of interaction is easy for human drivers (e.g., Schneemann & Gohl, 2016), but the way how automated driving technologies could communicate their intentions to pedestrians still remains as a question. A number of studies have examined how this communication could be enabled by technical means (e.g., Ackermann et al., 2019; Chang et al., 2017; de Clercq et al., 2019; Habibovic et al., 2018; Lee et al., 2019; Li et al., 2018; Mirnig et al., 2017). However, the communication should be efficient also to the other direction: the vehicle should be able to recognize the intentions of the pedestrians and other vulnerable road users (Rasouli and Tsotsos, 2019; Schwarting et al., 2019).

Based on the literature, problems automated driving technologies are encountering on roads are linked to safe, economical, and comfortable driving. Hence, what could be done to solve these problems?

2.3. Cognitive mimetics

Brown and Laurier (2017) as well as Chater et al. (2018) have concluded that human-computer interaction (HCI) and cognitive science could aid in designing better self-driving cars. One of the relevant paradigms could be cognitive mimetics (Kujala and Saariluoma, 2018; Saariluoma et al., 2018). This way of design thinking suggests that mimicking expert human drivers' information processing and thinking could be utilized for designing safer, more economical, and more comfortable automated driving technologies. The idea of design mimetics, that is, imitating physical and biological structures in nature for technology design, has been known since the fifties (Bar-Cohen, 2006). The core idea of cognitive mimetics is that instead of imitating these structures of nature, designers should focus on human experts' information processes and thinking when searching for model solutions (Kujala and Saariluoma, 2018; Saariluoma et al., 2018).

Thanks to its internal and information processing focus, cognitive mimetics differs from ethnographic approaches. Vinkhuyzen and Cefkin (2016), for example, studied how people behave in traffic and how this information could be utilized when developing and improving current automated driving technologies. They observed pedestrians and their behavior in order to teach automated driving technology to behave in "socially appropriate ways". However, in cognitive mimetics it is essential to pay attention also to the contents of experts' information processes (i.e., mental contents) (Newell and Simon, 1972). Recently, researchers in the field have started to realize the importance of examining human road users' behavior and its modeling in order to develop better automated driving technologies (e.g., Domeyer et al., 2019; Markkula et al., 2020; 2018; Merat et al., 2019). Due to the importance of anticipation of traffic events for successful driving (Stahl et al., 2013), the goal for investigating expert drivers' behavior here is to get a clearer idea of the information contents relevant in anticipatory driving.

2.4. Anticipatory driving and uncertainty

According to Pollatsek et al. (2006), novice drivers' fatality rate is eight times higher than the rate of highly experienced drivers. One causing factor is novice drivers' incapability to anticipate safety-relevant traffic events. Therefore, the anticipation of traffic situations is a critical component of driver competence, which allows drivers to maintain sufficient safety margins (Stahl et al., 2016). According to Tanida and Pöppel (2006), if the driving situation is perceived as familiar, drivers are able to anticipate what is going to happen next and to act accordingly. Conversely, if the driving situation is unfamiliar, drivers need to react to events. With human drivers, traffic flow, safety, and economical driving can be improved by moving from reactionary driving to anticipatory driving (Stahl et al., 2013).

Human experts' anticipatory skills (Clark, 2013) and the ability to focus processing situationally on task-relevant targets may be some of the key differences that separate human and machine intelligence. Based on neurological evidence, it has been proposed that the human brain is an advanced prediction machine (Clark, 2013). According to these accounts, its basic function is to continuously predict and anticipate the upcoming events and assess the uncertainty of the predictions. This framework of cognition stresses the importance of predictive uncertainty and its resolution in human attention allocation and behavior. In line with these ideas, it has been recently shown that experienced drivers' perceived uncertainty of upcoming traffic events on a freeway is a major factor in their visual information sampling (Kircher et al., 2019). In a similar vein, for instance, Meghjani et al. (2019) and Hubmann et al. (2018) stress the importance of modeling uncertainty in the development of decision-making for automated driving. Therefore, the analysis of the information contents of the expert drivers' anticipatory driving in this study was focused on uncertainties they recognize and resolve related to the unfolding traffic events.

3. STUDY 1 – EXPERT SAMPLE VALIDATION: HAZARD PREDICTION TEST

3.1. Method

In Study 1, the expert sample validation with a hazard prediction test was done in order to verify that the selected driving instructors are able to anticipate unfolding hazardous traffic events by a better rate than inexperienced or a mixed group of non-instructor drivers. While driving instructors teach their students, they anticipate possibly hazardous events and therefore are more prepared to act if it seems that the student driver cannot manage the situation. Furthermore, driving instructors are experienced in verbalizing their anticipations during driving lessons. Thus, we argue that driving instructors are well-trained experts in anticipating unfolding safety-relevant driving situations. In addition, the intention was to validate that the selected experts are able to anticipate such events that may currently be highly challenging for automated driving technologies.

3.1.1. Stimuli

Eighty driving clips, provided by a commercial UK company that provides hazard perception tests for learner drivers, were reviewed to select clips for the experiment. For evaluating the clips, knowledge of the previously reviewed challenges of the current automated driving technologies and the analyses by Hubmann et al. (2018), Lake et al. (2017), Lv et al. (2018), Rasouli and Tsotsos (2019), and Schwarting et al. (2019) on differences between human cognition and automated driving algorithms were utilized. Based on the evaluation, each selected clip was required to contain an unfolding hazardous event that human drivers should be able to anticipate – if they spot the relevant visual cue (s) – and which automated driving technologies perhaps would not be able to detect or anticipate. This could cause the automated driving technology to brake suddenly or even cause an accident.

Eventually, after reaching mutual understanding by two researchers, 28 out of 80 (35 %) clips were chosen that met the set requirements (see Table 1). The original clips were filmed in the UK and therefore were mirrored to respond to right-hand traffic, more familiar to Finland where the research was conducted. To transform the hazard *perception* clips into hazard *prediction* clips, each clip was edited to end to a black screen just after the hazardous event started to unfold, following the method by Crundall (2016), Jackson, Chapman, and Crundall (2009) and Ventsislavova et al. (2019). Effectively, each selected clip contained a situation that would potentially develop into hazardous event if neglected, such as a truck blocking driver's view, a ball flying over a street, or a street being too narrow for two cars to travel side by side. Hazard prediction test was chosen over hazard perception test since it can better discriminate between experts and novices (Crundall, 2016; Jackson et al., 2009).

3.1.2. Participants and experimental design

Participants were recruited via different mailing lists and by contacting driving schools directly. In total, 36 participants completed the experiment. The participants were divided into three groups: inexperienced (no driving experience, $n = 12$), mixed (varying driving experience, $n = 12$), and expert drivers (driving instructors, $n = 12$). The inexperienced group was included in order to test if hazard prediction ability comes with driving experience. A mixed group was included to represent large variation in cumulative driving experience, that is, to represent the driver population and to enable correlative analysis (experience vs. score). The driving instructor group was included to test if the formal training provides greater anticipation skills compared to a random sample from the driver population. Each participant had normal or corrected-to-normal vision. The demographics of the three participant groups can be seen in Table 2. It should be noted that the reported lifetime driving experiences in kilometers are estimations of the participants and they only include kilometers driven with cars. Since the kilometers are self-reported, the accuracy of estimations may vary between participants.

In order to study if the hazard prediction ability is a skill achieved with experience or if it can be rapidly learned with scenario-specific declarative knowledge (Rasmussen, 1983;1982), half of the participants in each group (inexperienced, mixed, and experts) were primed with generic written examples of possible hazards presented in the video clips (e.g., the door of the parked car suddenly opens). None or weak effect of briefing would stress the importance of using experts as the source of information in the subsequent study. Thus, the experimental design was 2 × 3 (briefing x group).

3.1.3. Materials and apparatus

The duration of the hazard prediction test clips varied in length from 4 to 43 seconds. Dell laptop computer with an external 22" screen was used to display the hazard prediction clips to the participants. The clips were presented in a randomized order with SMI Experiment Center 3.0 (SensoMotoric Instruments GmbH). SMI RED 500 remote binocular eye-tracking system (sampling rate 500 Hz) was utilized to track participants' eye movements (data not reported here). Sony HDR-XR500 video camera was used to record the participants' answers. IBM SPSS Statistics 24 was used for data analysis. The experimental setup is illustrated in Fig. 2.

3.1.4. Procedure

Upon arrival, participants read and signed the informed consent form. After that, participants were seated 60 cm from the screen. Before the actual experiment, each participant practiced with watching four videos which ended with a black screen just after the hazardous event started to unfold, similar to the actual videos, and answering to following questions after each clip: 1) What was the risk factor?, 2) What was the location of the risk factor?, 3) What happens next?, and 4) How would you proceed in the situation? Participants were instructed to evaluate the unfolding situation at the end of the clip and give an answer to each question as they feel is the correct answer regarding the unfolding situation. The same questions were asked for the videos in the actual experiment with the same instructions. If the answers were insufficient, instructions were repeated. Each participant group (inexperienced, mixed, and experts) received the same general instructions.

After the practice, the participants belonging in the briefing subgroups were told that similar hazards were repeating in the videos presented, and before the experiment started, they were given out a written hazard list to familiarize themselves with. The participants belonging to the no briefing group were instructed to look for a risk factor at the end of the video, with no information about the risk types or their recurrence.

In the actual experiment, participants watched 28 hazard prediction video clips in randomized order, which ended with a black screen just after the hazardous event started to unfold. After each video, participants were asked to answer four questions as they previously practiced. A small break after every 10 videos was offered to each participant. During the experiment, participants' oral answers were recorded with a video camera. The experiment took approximately 1.5 hours, and after the experiment, each participant received a gift card (15 €).

After the experiment, participants' verbal reports were analyzed and

Table 1
Clips for hazard prediction test.

Uncertainty	Situation(s) and environment(s) on parenthesis	Visual cue	Plausible difference in behavior – human driver vs. automated driving technology	# of clips
Is the road too narrow to accommodate the driver's own car and oncoming cars?	Narrowing road (Street)	Narrowing road ahead with oncoming traffic	Human driver is able to anticipate that the road ahead is too narrow for all vehicles and therefore waits for the oncoming vehicle to pass or adjusts speed or lane position. Automated driving technology could continue driving without decelerating, causing sudden emergency braking or an accident. Possible problem: lack of scene understanding (Lake et al., 2017).	6
Is there occluded traffic crossing or merging?	Poor visibility (Street)	Stationary objects occluding moving vehicles	Human driver is able to notice the moving vehicle behind the stationary object and anticipate that the vehicle might turn towards and therefore decelerates gently. Automated driving technology may not be able the recognize the moving vehicle behind the occluding object, causing sudden emergency braking or even an accident. Possible problem: occlusion and lack of intuitive physics (Hubmann et al., 2018; Lake et al., 2017).	5
Is there oncoming traffic behind the vehicle that is to be passed?	Passing, curvy road, poor visibility (Highway, Street)	Passing required with poor visibility ahead (e.g., curvy road ahead, truck blocking part of the road)	Human driver can realize that due to poor visibility of the road ahead it is not possible to see if there is oncoming traffic approaching behind the vehicle that is to be passed. Human driver is able to anticipate that oncoming traffic is a possible scenario and therefore slows down. Automated driving technology could continue driving with the same speed resulting in sudden emergency braking or an accident in case of oncoming traffic. Possible problem: lack of scene understanding and intuitive physics (Hubmann et al., 2018; Lake et al., 2017).	4
Are the vehicles on parallel side road going to merge in front?	End of a parallel road, traffic merging (Street)	Vehicles driving on ending parallel side road	Human driver is able to anticipate that the adjacent side road is ending and vehicles on the road may join the main road ahead and therefore adjusts speed accordingly. Automated driving technology may not detect an ending side road and anticipate that the traffic will soon merge, causing a possible sudden emergency braking or even an accident. Possible problem: lack of scene understanding and psychological reasoning (Lake et al., 2017).	3
Is the faster vehicle in front going to change lanes?	Slower traffic ahead, faster vehicle on the adjacent lane (Freeway)	Vehicles having speed differences on a two-lane road (e.g., faster vehicle approaching slower vehicle on the adjacent lane)	Human driver is able to anticipate that the faster vehicle may pass slower traffic ahead and move into the driver's lane. Due to the prediction, human driver is able to be prepared and adjust speed more gently than automated driving technology and even avoid an accident. Possible problems: lane change in heavy traffic (Lv et al., 2018) and lack of psychological reasoning (Lake et al., 2017).	2
Is the road/lane too narrow to accommodate driver's own car and oncoming vehicles that are passing?	Slower traffic ahead, passing (Freeway, Street)	Oncoming vehicles or vehicles in front passing slower traffic (e.g., cyclists, motorcyclists) far ahead	Human driver is able to anticipate that the road is too narrow for all the vehicles to travel parallel and is, therefore, able to slow down and/or adjust the lateral position of the car. Automated driving technology may not be able to detect the trajectory deviation of the passing car and is therefore unable to adjust its own speed and lane position causing possibly a sudden emergency braking or an accident. Possible problem: lack of scene understanding (Lake et al., 2017).	2
Is there someone stepping out of the parked car?	Parked cars (Street)	Parked car's door opening / person inside a parked car	The sudden opening of parked car's door or seeing a person inside a parked car are cues to a human driver that the door may open entirely and blocking the driveway. Due to the anticipation, human driver is able to be prepared, yield, and decelerate more gently in advance than automated driving technology might. Possible problems: driving too close to a parked car (Lv et al., 2018) as well as lack of scene understanding and psychological reasoning (Lake et al., 2017).	2
Is the pedestrian going to cross the street in front?	Pedestrian planning to cross the road (Street)	Pedestrians showing intentions to cross the road	Human driver is able to anticipate the trajectories of the pedestrians based on their behavior (e.g., looking both ways) or position even if they are stationary and, therefore, is cautious. Automated driving technology may not be able to recognize pedestrians' intentions since they are still on the walkway and not yet crossing. Sudden crossing could cause sudden emergency braking or an accident for automated driving technology. Possible problems: lack of understanding social cues (Rasouli and Tsotsos,	2

(*continued on next page*)

Table 1 (*continued*)

Uncertainty	Situation(s) and environment(s) on parenthesis	Visual cue	Plausible difference in behavior – human driver vs. automated driving technology	# of clips
			2019; Schwarting et al., 2019) and psychological reasoning (Lake et al., 2017).	
Is the occluded pedestrian going to re-appear behind the object and cross the road in front?	Occluded pedestrian (Street)	Stationary object occluding walking pedestrian momentarily	Human driver is able to anticipate that the pedestrian may re-appear behind the van and cross the road and therefore decelerate accordingly. Automated driving technology may not able the recognize the existence of the pedestrian behind the van, causing sudden emergency braking or accident. Possible problem: momentary occlusion and lack of intuitive physics (Hubmann et al., 2018; Lake et al., 2017).	1
Is emergency vehicle approaching and yielding needed?	Emergency vehicle approaching (Street)	Emergency vehicle's blue lights are approaching	Human driver is able to notice well in advance the blue lights and anticipate that yielding maneuvers may have to be made. For automated driving technology, the emergency vehicle may be detected as a crossing object to be avoided like any other, but it may not have an understanding that yielding well in advance is compulsory. Possible problem: identifying emergency situations (Lv et al., 2018).	1
Are kids going to run onto the street after the ball?	Kids playing (Street)	Ball flying over the street	Human driver is able to anticipate that a ball flying over the street indicates playing kids ahead and decelerates in advance accordingly. The ball does not indicate playing kids to automated driving technology, which might lead to sudden emergency braking or even an accident. Possible problem: lack of psychological reasoning (Lake et al., 2017).	1
Is the cyclist going to yield?	Cyclist in a traffic circle (Street)	Cyclist slowing down and approaching a traffic circle	The cyclist in the traffic circle tries to maintain momentum and, as a result, does not stop but signals the driver of yielding by entering the traffic circle with a slower speed. Human driver can anticipate the yielding of the cyclist. Since the cyclist is not stopping before entering the traffic circle, automated driving technology could perform sudden braking. Possible problems: lack of understanding social cues (Rasouli and Tsotsos, 2019; Schwarting et al., 2019), lack of psychological reasoning and intuitive physics (Lake et al., 2017).	1

Table 2
Demographics of the participant in hazard prediction test.

	Inexperienced group	Mixed group	Expert group
Age range	21–36	21–35	27–62
Mean age	M = 27.1, SD = 4.9	M = 27.1 SD = 4.8	M = 46.3, SD = 11.4
Gender	7 females, 4 males, 2 not disclosing gender	3 females, 9 males	3 females, 9 males
Range of driving experience in years	0	0.5 – 17	9.5 – 44
Mean driving experience in years	0	M = 8.4, SD = 5.2	M = 28.4, SD = 11.2
Range of self-estimated lifetime driving experience in kilometers	0 km	200 km – 1 000 000 km	280 000 km – 2 000 000 km
Mean self-estimated lifetime driving experience in kilometers	0 km	M = 202 475, SD = 331 317	M = 798 222, SD = 548 996

Fig. 2. The experimental setup.

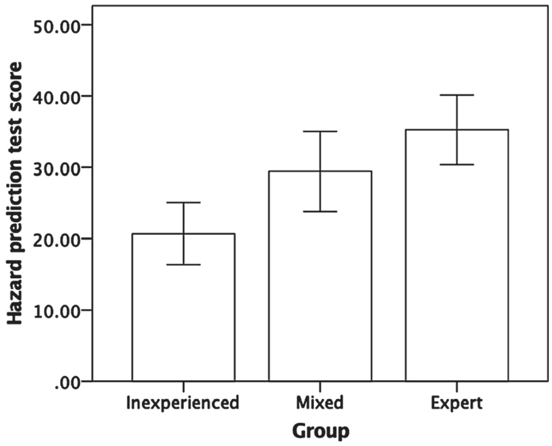

Fig. 3. Hazard prediction test score per group (mean, n = 12). Bars: 95% CI.

rated by two researchers. They were given 0–3 points for one video in total, depending on whether the participant had explicitly recognized the risk factor and its location and anticipated the development of the situation correctly. Correct answers were also accepted if the participant recognized several risk factors, of which one was the correct answer for the video. The fourth question (How would you proceed?) was not rated and the results are not reported here. Therefore, the maximum score for the videos was 84 points (3 × 28).

3.2. Results

Since the hazard prediction test scores and overall driving experience in kilometers were non-Gaussian, medians are reported here instead of means. The hazard prediction test scores (N = 36, interquartile range in parentheses) ranged from 12 to 48 points, and the overall median was 29 points (15). The median scores per group were: inexperienced 20.5 (11.0), mixed 30.0 (15.0), and experts 34.0 (10.0). The mean scores per group are illustrated in Fig. 3.

A factorial 2 × 3 ANOVA was conducted to investigate the interaction effects of briefing and group on hazard prediction test scores. There was no significant interaction between the factors (p = .490). Significant main effect of group was found on hazard prediction test scores: $F(2, 30) = 10.15$, $p = < .001$, $\eta_p^2 = .404$ (large effect). Due to the non-gaussian hazard prediction test score distribution in the inexperienced group, pairwise comparisons between groups were conducted with nonparametric tests. Similar to ANOVA, Kruskall-Wallis H test indicated that there were significant differences in hazard prediction test scores between the different groups, $\chi^2(2) = 13.632$, p = .001. According to Mann-Whitney U test, there were significant differences between novices and mixed group, mixed group scoring higher (U = 30.50, p = .016, d = 1.11 [large effect size]) and between novices and experts, experts scoring higher (U = 12.50, p = .001, d = 2.01 [large effect size]). The effect size of the difference between mixed group and experts was moderate (d = 0.71), but the difference was not significant with this sample size (U = 43.50, p = .099, n = 12).

For testing the association between lifetime driving experience in kilometers and hazard prediction test scores, Spearman's rank-order correlation was used. For the analysis, the group of novices was omitted since they do not have any driving experience, and therefore here N = 24. A moderate association between driving experience and hazard prediction test scores was found (see Fig. 4): ρ = .425, p = .038. However, even though the association between age and driving

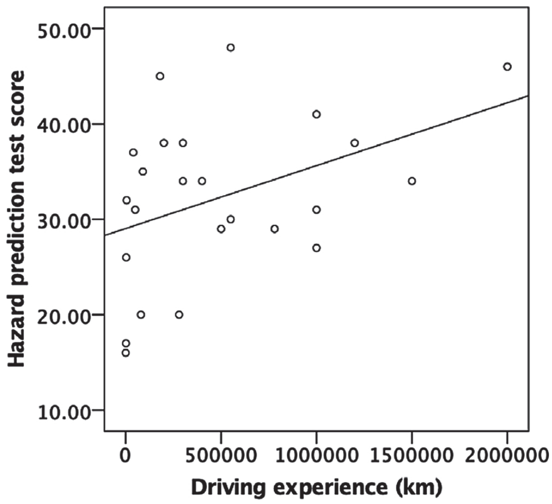

Fig. 4. Hazard prediction test score per lifetime driving experience (N = 24).

experience was strong ($\rho = .834$, $p < .001$), there was no significant association between age and hazard prediction scores ($\rho = .241$, $p = .256$).

3.3. Discussion

Based on the results, the driving instructors can be considered as experts compared to the inexperienced and also at least a subsample of them compared to the mixed group of non-instructor drivers. A moderate association between driving experience and hazard prediction test scores was found, which suggests driving experience can explain some of the variance in the scores. This is consistent with previous studies (e.g., Crundall, 2016; Jackson et al., 2009; Stahl et al., 2016; Ventsislavova & Crundall, 2018) and implies that the ability to anticipate traffic situations evolves with experience. It should be noted that the lifetime driving experience in kilometers was self-estimated and there could be differences between individuals how accurately they can estimate their experience. However, the correlation between age and experience and insignificant correlation between age and test score suggest general validity of the self-reported driving experience measurement, even if there might be inaccuracies in individual reports.

However, the driving experience does not alone explain the drivers' hazard prediction test scores. Further, since novices were also able to score in hazard prediction test above chance, anticipating hazardous situations is something that humans may be able to do at some level regardless of driving experience – perhaps with the help of intuitive psychology and intuitive physics, causal reasoning, and utilizing previous experiences in other domains (Lake et al., 2017). However, priming of participants with declarative knowledge about possible hazardous scenarios did not affect the scores. This may be due to a number of factors but suggests that the anticipation skills cannot be acquired based on written examples, at least in the short time provided to study the model scenarios. This finding stresses the importance of using domain experts as the source of situational information in studies such as our following on-road study.

4. STUDY 2 – FIELD STUDY WITH EXPERTS: ANTICIPATION IN REAL TRAFFIC WITH PROSPECTIVE THINKING-ALOUD

4.1. Method

In Study 2, six experts – validated with the hazard prediction test – took part in the field study where they were prospectively thinking aloud of unfolding traffic situations while driving on public roads. The purpose of the field study was to examine what driving-task relevant the experts are anticipating to happen.

4.1.1. Participants

The ages of the participants ranged from 36 to 56 ($M = 47.5$, $SD = 10.5$), lifetime driving experience from 280 000 to 2 000 000 kilometers ($M = 696\ 667$, $Mdn = 525\ 000$, $SD = 649\ 821$) and teaching experience from 5 to 38 years ($M = 10.8$, $Mdn = 9.0$, $SD = 14.1$). Their mean score in the hazard prediction test was 34.5 points ($SD = 10.7$). Three of them belonged to the no-briefing group and three of them to the briefing group in the validation experiment.

The ethical review board was inquired about the requirement of ethical approval and the study was allowed to take place after installation of a secondary brake pedal for the experimenter for the case of emergencies.

4.1.2. Materials and apparatus

The length of the predefined route was 57.2 kilometers (see Fig. 5) and driving the route took approximately one hour and ten minutes. The route was selected to contain a representative sample of local road environments: freeways (with controlled access), two-way highways, as well as suburban and city streets, and a parking lot.

Toyota Prius (2009) with an extra brake pedal was used in the experiment. For recording the road scene and thinking-aloud data, MoviePro application for iPhone 8 and an external microphone was used. Google Maps application, running on 10.5" iPad Pro, was used for providing route guidance (see Fig. 6). Speedometer application was placed next to the route guidance for enabling recording the GPS speed on the video. Participants' eye movements were recorded with head-mounted Ergoneers' Dikablis Professional eye-tracking system (data not reported here). Transcription of the thinking-aloud data was done using Noldus Observer XT 12 software.

4.1.3. Procedure

After informed consent and before the experiment started, each participant watched two training videos (1.40 minutes and 0.28 minutes) that were recorded on the same roads they were about to drive. While they watched the videos, they were instructed to anticipate aloud what driving-task relevant is going to happen next in the traffic and how it affects their behavior and maneuvering. As for the feedback during the training videos, we encouraged the participants to verbalize more actively the unfolding traffic situations, if necessary, which is typical for the thinking-aloud method.

After they were familiar with the prospective thinking-aloud method, they received the instructions for the drive. They were asked to obey traffic rules and follow the predefined route. During the drive they would be prospectively thinking aloud as they had practiced earlier. Theoretical models of multitasking performance, such as Wickens' (2008) Multiple Resources Theory, suggest that concurrent verbal-vocal tasks during visual-manual tasks (e.g., driving) do not interfere severely with each other. Further, according to Drews et al. (2008), when the topic of the conversation while driving is the surrounding traffic, it helps the driver to share situation awareness with the passenger. In our study, the experts were talking about the prevailing traffic and driving situations, and therefore we suggest that the think-aloud protocol did not distract them. In addition, the driving instructors are used to verbalizing driving situations to their students.

Further, Drews at el. (2008) propose that if the driving condition is demanding, the complexity of the conversation decreases. Before the on-road study, we gave instructions to our experts that if the driving situation needs their full attention, they can communicate it after the situation is under control. There were occasions where the drivers used this opportunity and the confronted uncertainty was communicated after the situation had cleared.

All participants completed the same route approximately at the same time in the afternoon close to rush hours in order to have more potential interactions with other traffic (with one exception: noon). The visibility during the trials was normal, although there was a light rain shower during the drives of two participants. After completing the route, the expert drivers received a gift card (15 €).

4.1.4. Data analysis

The prospective thinking-aloud data consists of six audio-visual recordings in real traffic. On average, one recording lasted for 1 hour and 10 minutes (time range from 65 to 75 minutes). The prospective thinking-aloud data from the videos were transcribed into textual format, resulting in 1277 utterances. The utterances were transcribed according to the start and the end of a comment. In addition, driving speed when the comment started and the speed when the comment ended were coded based on the GPS speed visible in the videos. On average, one participant produced 212 utterances (range: 80–408). In total, 124 utterances were excluded from the analysis, as their contents were not notions of uncertainty, such as "The pavement has been repaired a bit" and "That is good". Thus, individual notions of uncertainty were analyzed from a total of 1152 utterances. It should be noted that one utterance could include more than one uncertainty notion. Through this analysis, 1881 individual notions of uncertainty were listed. On average, one participant produced 313 individual notions of

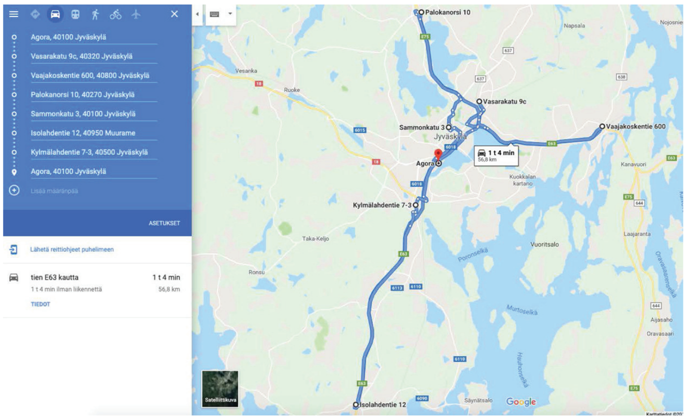

Fig. 5. The predefined route used in the study.

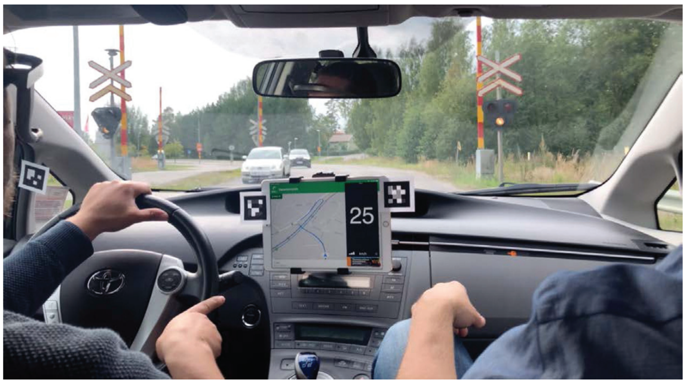

Fig. 6. Backseat view from a video.

uncertainty (range: 150–684).

A conceptual framework was developed to aid qualitative content analysis of the prospective thinking-aloud data. The framework provides a theoretical lens to guide the analysis, but not in a restricting or excluding manner. Thus, contents outside of the conceptual framework are analyzed with inductive content analysis (Mayring, 2000), if relevant to the research problem. A conceptual framework explicates the focus of the analysis by constructs, key factors or variables, and possibly the presumed relations among them (Miles and Huberman, 1994). Here, the focus was on examining driving-relevant uncertainties. Therefore, the conceptual framework consists of the following factors: uncertainty, environment, and goal.

The first factor, uncertainty, is defined as the unpredictability of a task-relevant event state (Clark, 2013) and is a central factor in the conceptual framework guiding the analysis. Contents of uncertainties were analyzed from the data with inductive content analysis (Mayring, 2000). The second factor, environment, consists of components that are defined according to the selected driving route. These components are freeway (with controlled access), highway (with two opposite lanes and crossings), street, traffic circle, and parking lot. The third factor of the framework, goal, consists of safety, economy, comfort, and wayfinding. Safety, economy, and comfort were selected based on the reviewed literature representing higher-level goals in driving and indicating these three as the main goals of improvement in automated driving technology. Wayfinding is an additional goal that is typical at the strategic level of driving (Matthews et al., 2001).

The conceptual framework also assumes relationships between the factors (Miles and Huberman, 1994). Here, the factors are related to each other in a context-dependent manner. The focus of the analysis was to understand what kind of uncertainties are expressed in what kind of contexts and to what kind of goals these uncertainties relate to.

Qualitative content analysis was conducted with the aid of the conceptual framework. Qualitative content analysis is utilized in systematic text analysis. The goal is to categorize information contents, usually with an inductive approach (Mayring, 2000). A central benefit in conducting inductive content analysis is its ability in detecting and developing categories with rich descriptions through analysis iterations of the information contents under analysis. In the iterative development process of the descriptive categories, overlapping categories were reanalyzed to be merged (Mayring, 2000). The analysis of the prospective thinking-aloud data for enabling cognitive mimetic design of automated driving solutions followed this procedure.

First, the environments were coded from the transcriptions and from the video recordings to ensure correct coding. All the videos were transcribed by two independent transcribers in sequential order, to diminish the possibility of rater factor to occur (see e.g., Gwet, 2014). Second, by further familiarization of the data, a category of situations was inductively developed as one independent entity. Situations were defined based on the transcribed sentences together with the synchronized videos as temporary issues and conditions occurring in specific environments. The descriptive subcategories of situations (e.g., traffic lights, road construction, traffic sign, exit ramp, intersection, congestion) were created through inductive content analysis. After this, two coders, who were present during the original drives, analyzed the data.

Coder 1 sorted all the data in Excel according to environment and situations in order to begin the context-dependent extraction of expressed uncertainties. If the uncertainty notions were difficult to understand due to the context-dependent sorting, the original sorting of the data by timestamp was displayed to ensure correct categorization of the uncertainty notions by the sentences preceding or following the sentence in question. The final uncertainty category consists of 83 descriptive subcategories that were written in the form of a question in order to illustrate the uncertainty related to the specific situation (cf. Table 1).

The category titled as action was developed to illustrate context-dependent uncertainty-related adaptive actions. The category of actions was analyzed by Coder 1 from the sentences if action-related notions were made. All notions did not include actions to be carried out. Uncertainty categories were iterated to develop a final set of subcategories to represent different uncertainty notions. After this, the data was further synthesized by analyzing goal (safety, economy, comfort, wayfinding) for each of the uncertainty subcategory. After Coder 1 had extracted the uncertainty notions and categorized these into subcategories, Coder 2 went through these coded uncertainties and possible divergent interpretations were discussed and resolved. After this, the number of notions and participants per notion for each uncertainty subcategory was calculated as presented in Tables 3 and 4. Finally, Coder 2 translated all the categories and selected example quotations into English and Coder 1 went through the translations to inspect that that there were no details lost during the translation process. Due to the nature of the process, we were not able evaluate inter-rater reliability numerically but these was a high level of agreement among the two coders in each step.

4.2. Results

The most important of the final uncertainty subcategories (83) that were related to safe, economical, and/or comfortable driving are listed in Tables 3 and 4. However, there were some subcategories that were excluded from the results. Six subcategories related to wayfinding (e.g., Are we on the route?; To which direction should we continue now?) were found – these are omitted from the results, as navigation should not be a challenge to current automated driving technologies. Some of the uncertainties are highly relevant for a human driver (e.g., What is the current speed of the car?; Is someone approaching in the blind spot?) but irrelevant for automated driving technology and were left out from the data. The rest of the excluded subcategories are relevant also for automated driving technologies but should be easily resolved by the current level of technology (e.g., Are there oncoming cars or pedestrians ahead?; What is the speed limit?; Are there cars beside or behind our car?; Is it slippery?). All of these mentioned uncertainties (20) were excluded from this report. These excluded uncertainties include the only uncertainty notion that was related to the operational level of control of the vehicle (i.e., Can I hold the control of the car?) – all the other uncertainties were related to tactical or strategic level situation awareness (Matthews et al., 2001).

First, Table 3 presents uncertainties (34, 54%) of other road users' behaviors, awareness and/or intentions. Then, Table 4 lists all the uncertainties (29, 46%) that are not (directly) related to behaviors, awareness or intentions of other road users. Both tables also include those uncertainties that were related only to a few notions, as these may be important even if the traffic conditions did not lead to these kinds of situations for all the participants. The Goal category can relate to one's own as well as other road users' safety, comfort and/or economy.

4.3. Discussion

Most of the expressed uncertainties related directly to the behaviors, awareness, or intentions of other road users (54%, Table 3). These uncertainties include, for instance, recognizing other road users' intentions, signaling own intentions to them, other road users' lane changing actions, and other's situation awareness (e.g., Is other traffic keeping safe following distance?; Is other road users' visibility sufficient?). A notable uncertainty in Table 3 is related to the possibility – or even expectation – that others will not obey traffic rules. Central actions related to resolving these uncertainties are deceleration (slowly), giving way to others, increasing following distance, giving turn signals well ahead, and eye contact. Some of the uncertainties related to social behavior in traffic that have also been raised in previous studies (e.g., Brown & Laurier, 2017; Chater et al., 2018; Mahadevan et al., 2018; Rasouli & Tsotsos, 2019; Vinkhuyzen & Cefkin, 2016).

However, there were also a number of other types of uncertainties that are not, at least directly, related to the behaviors of other road users (46%, Table 4). Many of these uncertainties relate to one's own behavior, for example, what is the proper speed and safe deceleration/acceleration rate, is the headway distance sufficient, when there is a sufficient gap to merge or change lanes, and is the alignment of the car on the lane such that there is sufficient space for other road users (if needed). Some relate to one's own situation awareness, such as visibility ahead, and interactions with traffic lights and other cars standing in these. Central actions to resolve these uncertainties were similar to those in Table 3, for instance, keeping distance, driving with a steady speed, decelerating slowly, accelerating slowly or with force when required, and alignment.

Whereas one could argue that many of the uncertainties in Table 4 are due, in the end, also to the necessity to interact with other road users, these are more related to uncertainties of what are the optimal ways to control one's own vehicle in the arising situation to increase safety, fluency, comfort, and economy in the traffic system, instead of uncertainties related directly to social behaviors. It is worthwhile to stress that the uncertainties expressed by the human expert drivers did not only focus on ensuring one's own, but also other road users' safety, comfort, and economy. Most of the listed goals were related to safety. There are only six uncertainties that are not directly related to safety, and only one of these is under the uncertainties related to social behaviors in Table 3.

The method of the field study was similar to the methodology used by Kircher and Ahlström (2018). However, the emphasis was here on the prospective thinking-aloud and the focus of analysis on the anticipations and uncertainties of the expert drivers, as justified in related literature, whereas Kircher and Ahlström (2018) focused on evaluating the utility of various methods to assess driver's attentional state. They found and stressed the importance of taking into account the intentions of the driver for this analysis, and that thinking-aloud was an appropriate tool to gain insight into the driver's actual situational mental representations. Kircher and Ahlström (2017; 2018) argue that there is currently not enough a priori understanding of the minimum attentional requirements for safe driving applicable to any given driving situation. They suggest that prototypical situations and maneuvers in traffic as well as the situationally relevant information targets and agents in these should be defined to accumulate this understanding. The prospective thinking-aloud method seems to serve also for this purpose, and in particular, of studying the minimum information requirements related to the sufficient Level 3 situation awareness in real-world scenarios (Endsley, 1995).

Most of the found uncertainties relate to dynamic and temporal goal-relevant variabilities in the driving situation, and in particular the ones related to interactions with other traffic. There are only a few uncertainties that are spatial and/or more static by nature, such as those related to prevailing speed limit, nature of the intersections (equal or not), holes or objects on the road, traffic lights, optimal driving lines, lengths of entrance ramps, and the sufficiency of space on the road to fit passing vehicles. The dynamic uncertainties represent time-critical information requirements in driving whereas the static uncertainties are related to information requirements of the infrastructure of the traffic system (Kircher and Ahlström, 2017). The data suggests that most of these requirements in driving may be dynamic

Table 3
Uncertainties of other road users' behaviors, awareness and/or intentions (sorted by number of notions, *n*).

Uncertainty	Situation(s)	Environment(s)	Goal	Expert's example quotation indicating uncertainty	Action(s)	Notions (n)
Are other road users aware of my intentions?	Turn, lane change, exit ramp, merge, traffic circle exit, parking	Street, traffic circle, freeway, highway, parking lot	Safety, comfort	"Slowly we drop our speed and now turn signal on to indicate that we are turning – let's tell that to others too. Well, well, there is a friendly co-driver, and this is how it goes."	Use of turn signal in advance, alignment, soft deceleration, creeping, way giving, turn signal off, eye contact	101 (6)
Is there going to be a rear-end collision?	Traffic lights, intersection, exit ramp, driving in traffic queue, bus travelling behind, animal transport travelling behind	Freeway, street, traffic circle, highway	Safety	"That red car behind us is quite close. But no worries, we are keeping good following distance and we are not going to do anything sudden."	Distance, decelerating slowly, checking rear-view mirror, avoiding sudden braking, anticipatory deceleration, alignment, use of turn signal in advance	78 (6)
Who is obligated to yield?	Traffic lights, 4-way intersection, intersection with yield sign, crosswalk, turning, stationary bus, entrance ramp	Street, traffic circle	Safety	"Arrow light, there shouldn't be anyone we should yield. Then equal crossroads – cars approaching from the right-hand side and from ahead, but those are so far that they will not disturb us."		69 (6)
Are others obeying traffic rules?	Traffic lights, driving order in intersections, obligation to yield, turn, crosswalk, entrance ramp, speed limit, pedestrians not using crosswalk for crossing, drivers don't know how to drive in a two-lane traffic circle	Street, traffic circle, freeway, highway	Safety	"We have a green arrow light, therefore there shouldn't be any pedestrians ahead but still there is a chance that someone comes behind the yield sign."	Deceleration, way giving, eye contact, yielding, being adaptable, checking rear-view mirror, distance, creeping, pulling away slowly in traffic lights	63 (6)
Is the vehicle's following distance sufficient (behind us)?	Driving in traffic queue	Freeway	Safety	"The car behind us is not keeping safe following distance to us and that's why we need to keep extra distance to the car in front of us."	Checking rear-view mirror, avoiding sudden movements, keeping steady speed	25 (6)
Do other drivers need a lane change?	Lane change, entrance ramp	Street, freeway	Safety, comfort	"Then we are reducing the speed and keep distance to the car in front of us if someone on the parallel lane wants to change lanes."	Deceleration, distance, way giving	15 (5)
How is the traffic far ahead on the route?	Exit ramp, end of freeway, curve, congestion, entrance ramp	Traffic circle, freeway, highway, street	Safety, economy	"At this time, there must be a traffic congestion."	Deceleration	13 (6)
Is parked car's door going to open?	Road construction, rain, parked cars, narrow streets	Street, parking lot	Safety	"We shouldn't drive (too close to the parked cars) in a way that our side mirrors bang – there is a chance that the parked car's door will open suddenly."	Distance between own and parked cars	11 (4)
Is other traffic keeping safe following distances?	Driving in traffic queue	Highway, freeway	Safety	"I can see brake lights. That black car is too close to that other car."	Distance, steady speed	10 (5)
What are the intentions of other road users?	Turn, lane change, merge, intersection, pedestrian	Street, highway	Safety, comfort	"Interesting to see what that white car is planning to do."	Deceleration	10 (4)
Do other drivers need a last minute's lane change?	End of lane, end of freeway	Highway, freeway	Safety, comfort	"The parallel lane is ending so we should observe the left-hand side mirror in case there is still someone rushing to our lane."	Checking rear-view mirror	9 (3)
Are other vehicles able to enter the freeway?	Entrance ramp	Freeway	Safety, comfort, economy	"The main thing is not to brake suddenly in these kind of situations [others entering the freeway] but releasing throttling usually helps co-drivers to accelerate enough."	Way giving, slowing down, checking mirrors, distance	9 (3)
Is there an emergency vehicle approaching?	Chance of emergency vehicles	Street, parking lot, highway	Safety	"There's green arrow light but still it's good to check that there are no emergency vehicles approaching."	Checking rear-view mirror	9 (3)
How is the traffic queue moving?	Driving in traffic queue, traffic lights, intersection, heavy traffic	Street, highway	Safety, comfort, economy	"Traffic lights change to green but there's a big truck in front, so the queue is not moving fast. That means that there's no rush to move forward and I'll show that to my co-drivers too.	Looking far ahead, checking rear-view mirror, use of turn signal, driving in neutral gear	8 (3)
Is there a train approaching?	Crossing train tracks	Street, railroad crossing	Safety	"Everyone should remember to pay attention when crossing train tracks. You never know what happens there."	Deceleration	7 (6)
Is oncoming traffic passing (two lanes)?	Passing	Highway	Safety, comfort	"This road is changing into two-lane road. It means that now we have to observe oncoming traffic and their possible passing."	Deceleration, way giving	6 (4)
	End of lane, intersection, traffic lights	Street			Deceleration, distance	6 (3)

(*continued on next page*)

Table 3 (*continued*)

Uncertainty	Situation(s)	Environment(s)	Goal	Expert's example quotation indicating uncertainty	Action(s)	Notions (n)
Is someone in front going to change lanes?			Safety, comfort	"Then the parallel lane is ending but no one is there."		
Is the traffic queue moving?	Traffic lights	Street	Comfort	"Let's make sure that the queue is moving nicely and no one's engine shuts off or something like that."	Driving in neutral gear, slow speed	6 (2)
Why is the traffic queue slowing down/ stopping/ congested?	Driving in traffic queue, intersection, oncoming traffic	Street, highway	Safety, comfort	"I bet there is a big truck ahead making a wide turn."	Deceleration	5 (3)
Are there children on sidewalk/behind parked cars?	Turning, children crossing sign, daycare nearby, parked cars	Street	Safety	"Well, are there small pedestrians behind the car?"		5 (3)
Can others see my vehicle?	Rain, distracted drivers, pedestrian crossing street	Highway, parking lot, street	Safety	"I'll check that the headlights are on because I'm not sure if they work automatically. Due to rain, I want to make sure that I'm visible to others."	Head lights, eye contact	4 (4)
Are we approaching an intersection at the same time (merging)?	Merging, curve, end of lane	Freeway	Safety, comfort	"There's a curve ahead, we are not yet going to accelerate and we need to keep safe following distance so we all aren't in the intersection at the same time – you never know what kind of drivers there are going to be."	Distance, monitoring other traffic	4 (3)
Is other road users' visibility sufficient?	Large vehicles	Freeway, highway	Safety, comfort	"There is a car approaching on the entrance ramp. Since it's a van, it can't necessarily see us."	Distance, deceleration	4 (2)
Is the car in front rolling to my car's nose?	Traffic lights	Street, highway	Safety	"If we drive close to the car that is in front of us, there's always the risk that it will roll to our nose."	Distance	3 (3)
Is the passing vehicle able to pass?	Passing	Freeway	Safety, comfort	"There's a car passing us, let's release the throttle a bit and it will be able to pass us smoothly."	Deceleration, checking mirrors	3 (2)
Is someone cutting in?	Exit ramp	Freeway	Safety, comfort	"Now someone started passing behind us. But it doesn't bother us because I don't think that it will cut in."		3 (2)
Is the bus merging in front of my car?	Bus merging from bus stop	Highway	Safety	"Okay, that bus is merging from the bus stop, well, and there it is – in front of us."		2 (2)
Is the truck with a trailer able to change lanes?	Lane change of truck with trailer, short entrance ramp	Freeway	Safety, comfort	"There is a green truck with a trailer in the intersection. Let's see when it will join the traffic."		2 (2)
Are there vehicles turning behind the bus?	Intersection, turning	Street	Safety	"Well, it seems that no one is approaching behind the bus, I can turn now."		2 (1)
Is someone approaching behind parked cars?	Parked cars, on-street parking	Street	Safety	"There are parked cars on the right-hand side, let's pay attention if someone is coming behind them."		1 (1)
Is a truck going to cause rear-end collision behind?	Line and truck behind keeping short following distances	Freeway	Safety	"I'm checking the queue behind, there's a truck really close to other cars. Now it's really important that we don't brake suddenly because the queue behind is so tight – so to speak."	Avoiding sudden braking	1 (1)
Is someone passing in front of my car in the same lane?	Driving in traffic queue	Highway	Safety, comfort	"If someone in front starts passing, we have to act along if it looks that the passing car is unable to return to its own lane in time – we will not start to compete, we rather slow down and help."	Deceleration, way giving	1 (1)
Are other vehicles able to change lanes safely?	Entrance ramp	Freeway	Safety, comfort	"Here we have to pay attention to traffic merging from right – sometimes you see quite interesting lane changes there."		1 (1)
Will driver behind notice if her/his car starts to nose?	Traffic lights, driver behind is reaching something from the floor	Street	Safety	"The driver behind us has lost something and is trying to find it from the floor. Hope he has hand brake on and won't nose and cause a head-on collision."		1 (1)

Table 4
Uncertainties not (directly) related to the behaviors, awareness, or intentions of other road users (sorted by the number of notions, *n*).

Uncertainty	Situation(s)	Environment(s)	Goal	Expert's example quotation indicating uncertainty	Action(s)	Notions (n)
What is the proper approach speed?	Curve, exit ramp, entrance ramp, traffic lights, speed limit signs, keeping safe following distance, turning, acceleration, traffic queue ahead, speed limit change ahead	Street, traffic circle, freeway, highway, parking lot	Safety, comfort, economy	"Traffic circle ahead, I'll let the speed drop slowly, 20 km per hour is an optimal speed in this traffic circle."	Deceleration, driving in neutral gear, Accelerating, increasing following distance, forcing others to decelerate near crosswalks, way giving, checking rear-view mirror, avoiding sudden braking	120 (6)
Is the following distance sufficient?	Traffic lights, speed limit, congestion, rain, passing, entrance ramp, entrance ramp (others), road construction	Street, freeway, highway	Safety	"I will release the throttle a little bit – in case that car will merge in front of us. That way we will keep safe following distance."	Distance, checking mirrors, deceleration, way giving	69 (6)
What is a safe gap to merge?	Intersection, entrance ramp, traffic circle entrance	Street, traffic circle, freeway	Safety	"And then there is a good gap for us to merge, yes that yellow car is so far away that it is save to merge."	Use of turn signal, deceleration, driving in neutral gear, alignment, way giving, speed adjusting, checking blind spot	67 (6)
Is stopping in the traffic lights needed?	Traffic lights	Street, highway	Safety, comfort, economy	"And the journey continues, those traffic lights seem to be red so there's no rush to get there just to wait in traffic lights."	Driving in neutral gear, deceleration, checking traffic lights of crossing traffic, checking rear-view mirror	61 (6)
When is changing lanes safe?	Lane change, traffic lights, entrance ramp	Street, traffic circle, highway, freeway	Safety	"There comes a van, I have turn signal on – then it's our turn to merge."	Way giving, use of turn signal, checking mirrors, checking rear-view mirror, checking blind spot, deceleration, accelerating, strong accelerating	47 (6)
Is the visibility sufficient?	Turning, intersection, crosswalk, exit ramp, other cars in entrance ramp, weather, traffic in front, rain	Street, traffic circle, freeway, highway, parking lot	Safety	"In this situation, let's devote to keeping sufficient following distance. There is a big truck in front of us, we don't want to drive too close to the truck since it makes it hard to see what is happening in front of it."	Distance, deceleration, creeping, distance, driving slowly	45 (6)
Is stopping needed?	Turning (car in front), braking (car in front), congestion, congestion in traffic circle	Street, traffic circle, highway, freeway	Safety, comfort, economy	"That car in front of us is going to turn. So, let's release the throttle."	Distance, deceleration, driving in neutral gear	35 (6)
Is the road/lane too narrow?	Narrow street, parked cars, turning on highway, alignment	Street, highway, parking lot	Safety	"Again, the street is narrow and there is an oncoming car. Let's wait here until the car passes us and then we have more space to turn."	Deceleration, alignment, way giving, checking mirrors	25 (6)
Are the road constructions causing exceptions in traffic arrangements?	Construction work sign	Street, freeway	Safety	"Construction work zone. There are no contemporary speed limit signs but still it's reasonable to be cautious if there are construction workers present."	Alignment to the right, way giving, deceleration, driving in neutral gear, changing lanes, use of turn signal in advance	23 (6)
Is braking needed?	Slower traffic ahead, exit ramp, road construction, moose warning	Highway, freeway	Safety, comfort, economy	"We are catching up the car in front of us. Let's release throttle and see to what speed that car is adjusting to."	Checking shoulders, deceleration, checking rear-view mirror, passing, engine braking	17 (5)
When can I start moving again?	Traffic lights, intersection, queue	Street, highway	Comfort, economy	"And slowly the queue starts to move forward."	Checking traffic lights of crossing traffic, 'sliding'	14 (4)
When are the traffic lights going to change?	Traffic lights, intersection, driving order	Street, highway	Comfort	"The traffic lights on left changed to yellow. That indicates that ours are going to change to green soon. And that's exactly what happened."	Deceleration, checking traffic lights of crossing traffic, 'sliding'	13 (4)
Is there time to cross the intersection before the traffic lights change into red?	Traffic lights	Street	Safety, comfort	"I will pay attention to traffic lights and check if we have time to turn before the light turns into red, I'm not just following the traffic."	Creeping, way giving	13 (4)
Are the traffic lights going to change?	Approaching traffic lights	Street, highway	Safety, comfort	"Is the light going to change to red? No."	Deceleration	11 (5)
Is braking needed in curve?	Exit ramp, entrance ramp	Street, freeway	Safety, comfort, economy	"The curve is quite sharp before the acceleration lane and that's why we shouldn't drive too fast."	Deceleration, driving in neutral gear	11 (4)
What is the proper pace to slow down?	Driving in traffic queue, speed limit sign, traffic lights	Street, freeway, highway	Comfort, economy	"Slowly we brake and that way we don't scare drivers in front of us or behind us."	Deceleration, driving in neutral gear, checking rear-view mirror, deceleration efficiently but safely	9 (2)

(*continued on next page*)

Table 4 (*continued*)

Uncertainty	Situation(s)	Environment(s)	Goal	Expert's example quotation indicating uncertainty	Action(s)	Notions (n)
Is passing needed?	Slower traffic ahead, 100 km/h speed limit	Freeway	Safety, comfort, economy	"There's going to be 80 km per hour speed limit soon so it's not reasonable to pass that guy – when our cars are side by side, the 80 km per hour zone starts."	Deceleration	8 (4)
Do I hit a pothole/object on the road?	Pothole on pavement, (small) object on the road, roadkill	Street, highway	Comfort	"Let's avoid those potholes on the pavement – that way it's more comfortable to drive."	Yielding, continue driving without yielding	7 (4)
When is the proper time to slow down?	Entering exit ramp	Freeway, highway	Safety, comfort, economy	"The car behind is really close to us. That's why I'm not going to brake strongly, more like throttling back so there is no surprise to the driver behind."	Use of turn signal in advance, checking rear-view mirror, deceleration at the exit ramp	7 (2)
If needed, is there space to yield to the adjacent lane?	Traffic merging, entrance ramp	Freeway	Safety, comfort	"There's an entrance ramp so it's good to check the left-side mirror and make sure that there is not a tight situation forcing us to yield – that's the last straw."	Checking left mirror, staying on own lane, way giving	5 (3)
Do I block the sidewalk?	Intersection, crosswalk	Street	Safety, comfort	"There is still enough space behind the car for pedestrians to cross."	Distance, alignment	4 (3)
Do I cause a rear-end collision?	Car in front, traffic lights	Street	Safety, comfort	"There's turning car in front, let's give proper time for the driver to turn."	Distance	4 (1)
Is sudden braking needed?	Driving in traffic queue	Traffic circle, highway	Safety, comfort, economy	"We have time to steadily slow down before the traffic circle and then we observe if sudden braking is needed."	Deceleration, use of turn signal, distance	3 (2)
Where can I park (safely)?	Parking	Parking lot	Safety	"In here the risks are these other cars starting off – at least there, that one on the left. And then we reverse."	Deceleration	2 (1)
Are there equal intersections ahead?	Speed limit sign, intersection	Street	Safety	"40 km per hour zone starts here, which indicates that there could be equal intersections ahead."		2 (1)
Is there enough space to pass?	Traffic lights	Highway	Safety, comfort	"There should be proper distance between us and the car in front of us – if that car breaks down, we can still pass by going around the car and we don't have to reverse."	Distance, deceleration	2 (1)
What is the safest driving line?	Traffic queue	Freeway, highway	Safety	"And certainly, we should drive [on the lane] as right as possible, and actually – there is a mention in the law that one should drive as right as possible."	Alignment	2 (1)
Is the entrance ramp long enough for merging?	Merging	Freeway	Safety, comfort	"No one is coming, then the entrance ramp – which is short – and then we speed up strongly and check again that no one is coming."	Accelerating with force	1 (1)
Is it possible to drive uphill at a steady speed?	Hilly road	Highway	Comfort, economy	"Now we start to accelerate because there is an uphill and we don't want our speed to drop."	Accelerating in advance before uphill	1 (1)

and time-critical, which is understandable for a dynamic visual-spatial tracking task.

From a methodical perspective, it is important to notice that anticipation is always a mental content. It is possible to anticipate only if people are able to represent the present and the possible future states of the situation in their minds. Chess players, for example, simulate possible future states in their minds when they search for the best move (Saariluoma, 1995). They generate and relate moves that are not present in their perceptual field, and thus, they can anticipate the possible courses of actions in their mental representations. The contents of their thoughts explain why they can represent future state of affairs and anticipate what will happen. Similarly, in the presented thinking-aloud protocol drivers generate future state of affairs and anticipate possible future courses of actions and the uncertainties of these. They mentally simulate what can happen and how they should act in a given situation to avoid negative outcomes of actions. The human ability to represent mental contents – that are often conceptual or qualitative – is decisive for the human way of anticipating possible future courses of events and to adapt their present actions to avoid accidents. How to enable this kind of generic capacity for automated vehicles without introducing computationally heavy world models is a challenging question. However, it seems that in order to improve the safety of automated driving to a human expert level – or beyond, this capacity is a requirement.

4.3.1. Limitations and future work

According to Lake et al. (2017), due to nonexistent world models, automated driving technologies cannot raise, for instance, the uncertainty about recognizing other road users' intentions or sufficient visibility. Based on the assumption of non-existing world models, the reviewed literature and the information publicly available online (e.g., https://www.tesla.com), some of the uncertainties in Tables 3 and 4 may be impossible to be recognized by current automated driving technologies. These uncertainties may remain out of reach of automated driving technologies for the distant future unless there are major advances made towards general artificial intelligence (Kujala and Saariluoma, 2018). However, the details of the state-of-the-art and developing commercial technologies outside academic knowledge are hard to find due to trade secrets. The authors are not experts in the engineering of automated driving technologies, and will not speculate which of the found uncertainties could or could not be recognized and/or resolved with current technology. We will leave this analysis for the domain experts and as a topic for further research. However, we believe this data is valuable for the developers in assessing the limitations of current state-of-the-art technology and in finding ways to improve situation awareness of future automated driving solutions. The introduced method and produced data can be utilized also for making automated vehicles to recognize such upcoming situations, in which the human should take over the vehicle, to enable timely take-over requests before safety-critical situations realize (Hecker et al., 2018).

On the other hand, many of the found uncertainties are probably not recognized by current automated driving technologies but could perhaps be recognized and resolved by the existing technologies. From a mimetic design perspective (Kujala and Saariluoma, 2018; Saariluoma et al., 2018), these are the most interesting uncertainties. With improved map data (e.g., Are road constructions causing exceptions in traffic arrangements?), machine vision (e.g., Do I hit the pothole/object on the road?), vehicle-to-vehicle and vehicle-to-infrastructure communications (e.g., Are traffic lights going to change?), and data fusion (e.g., If needed, is there space to yield to the adjacent lane?), many of these anticipatory capacities may possibly be implemented in today's or tomorrow's automated driving technologies. Further research should take each one of these uncertainties and create means for automated driving technologies to recognize and resolve these – if not yet being implemented.

The sample of driving instructors was quite small, although it seemed there was a saturation of data for the selected routes. As there was no direct control over the traffic conditions that is possible in a driving simulator, some of the situations were rare but still safety-relevant. With a larger sample, more of these events and possibly also other types of uncertainties could have been observed. Intuitively, all the found uncertainties seem to be such that these could be relevant across various traffic environments and cultures. However, the route was relatively short (57.2 km) and represented only the uncertainties relevant in the selected local traffic conditions and time of day, and therefore uncertainties relevant in other traffic environments, conditions and times of day (e.g., night) could be missing. In further research, the method should be applied to various traffic environments and cultures in order to reveal all the possible relevant uncertainty subcategories that are not handled by automated driving solutions.

In future studies, utilizing eye-tracking and vehicle data together with the prospective thinking-aloud method could enable more detailed quantitative analyses of the adaptive actions to the expressed uncertainties, such as speed and headway adaptations (cf. Kircher and Ahlström, 2018). This level of analysis might enable computational models of human expert drivers' decision-making and adaptations in situations with the recognized uncertainties (cf. Hubmann et al., 2018; Meghjani et al., 2019; Portouli et al., 2019) that may be useful for implementing these in automated driving algorithms.

5. CONCLUSIONS

We have introduced the prospective thinking-aloud method for analyzing how expert drivers (driving instructors) think as well as the anticipations and uncertainties of expert drivers related to safe, economical, and comfortable driving. The expertise of the driving instructors was validated with a hazard prediction test. As expected, driving instructors were able to anticipate unfolding hazardous traffic situations by a better rate than the other participant groups and it seems that this prediction ability evolves with practice.

The results of the field study indicate that there may be uncertainties in traffic that are perhaps not recognized or resolved with current automated driving technology solutions. It remains unclear if a great number of training data and great processing power are sufficient for overcoming these challenges. If the ultimate goal of this development is to create a fully autonomous vehicle that can cope in any complex driving situation with human road users, especially the social side of automated driving should be better understood.

However, the method also revealed a number of significant uncertainties that may not be considered in the development of automated driving technologies, but which may be recognized and resolved with existing technologies. Further, the introduced method may serve in enabling automated driving technologies to predict its probable failure, in order to alert the driver to take control well ahead of the failure (Hecker et al., 2018).

These findings and methodical contributions can be utilized when studying expert drivers' anticipations in different contexts, prototypical traffic situations and maneuvers and their information requirements for safe driving (Kircher and Ahlström, 2017; 2018), and for developing better automated driving technology by indicating automated vehicles' potential limitations as compared to expert human drivers.

CRediT authorship contribution statement

Hilkka Grahn: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing - original draft. **Tuomo Kujala:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing - review & editing, Supervision, Project administration. **Johanna Silvennoinen:** Methodology, Validation, Writing - review & editing. **Aino Leppänen:** Investigation, Writing - review & editing. **Pertti Saariluoma:** Conceptualization, Methodology, Writing - review & editing, Funding acquisition.

Acknowledgments

The research was partly funded by the project Ethical AI for the Governance of the Society (ETAIROS), funded by Academy of Finland. We thank Henrik Lassila for assistance with the experiments.

References

Ackermann, C., Beggiato, M., Schubert, S., Krems, J.F., 2019. An experimental study to investigate design and assessment criteria: What is important for communication between pedestrians and automated vehicles? Appl. Ergon. https://doi.org/10.1016/j.apergo.2018.11.002.
Bar-Cohen, Y., 2006. Biomimetics - Using nature to inspire human innovation. Bioinspiration and Biomimetics. https://doi.org/10.1088/1748-3182/1/1/P01.
Baumann, M., Krems, J.F., 2007. Situation awareness and driving: A cognitive model. Modelling Driver Behaviour in Automotive Environments: Critical Issues in Driver Interactions with Intelligent Transport Systems. pp. 253–265. https://doi.org/10.1007/978-1-84628-618-6_14.
Brown, B., 2017. The social life of autonomous cars. Computer (Long. Beach. Calif). 50 (2), 92–96. https://doi.org/10.1109/MC.2017.59.
Brown, B., Laurier, E., 2017. The trouble with autopilots: Assisted and autonomous driving on the social road. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI' 17 416–429. https://doi.org/10.1145/3025453.3025462.
Chae, H., Lee, M., Yi, K., 2017. Probabilistic prediction based automated driving motion planning algorithm for lane change. International Conference on Control, Automation and Systems. https://doi.org/10.23919/ICCAS.2017.8204250.
Chang, C.M., Toda, K., Sakamoto, D., Igarashi, T., 2017. Eyes on a car: An interface design for communication between an autonomous car and a pedestrian. Proceedings of 9th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications 65–73. https://doi.org/10.1145/3122986.3122989.
Chater, N., Misyak, J., Watson, D., Griffiths, N., Mouzakitis, A., 2018. Negotiating the traffic: Can cognitive science help make autonomous vehicles a reality? Trends Cogn. Sci. https://doi.org/10.1016/j.tics.2017.11.008.
Clark, A., 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav. Brain Sci. 36 (3), 181–204. https://doi.org/10.1017/S0140525X12000477.
Crundall, D., 2016. Hazard prediction discriminates between novice and experienced drivers. Accid. Anal. Prev. 86, 47–58. https://doi.org/10.1016/j.aap.2015.10.006.
de Clercq, K., Dietrich, A., Núñez Velasco, J.P., de Winter, J., Happee, R., 2019. External human-machine interfaces on automated vehicles: Effects on pedestrian crossing decisions. Hum. Factors. https://doi.org/10.1177/0018720819836343.
Dikmen, M., Burns, C.M., 2016. Autonomous driving in the real world: Experiences with Tesla Autopilot and Summon. Proc. 8th Int. Conf. Automot. User Interfaces Interact. Veh. Appl 225–228. https://doi.org/10.1145/3003715.3005465.
Domeyer, J., Dinparastdjadid, A., Lee, J.D., Douglas, G., Alsaid, A., Price, M., 2019. Proxemics and kinesics in automated vehicle–pedestrian communication: Representing ethnographic observations. Transp. Res. Rec. 2673 (10), 70–81. https://doi.org/10.1177/0361198119848413.
Dong, C., Chen, Y., Dolan, J.M., 2019. Interactive trajectory prediction for autonomous driving via recurrent meta induction neural network. Proceedings - IEEE International Conference on Robotics and Automation. https://doi.org/10.1109/ICRA.2019.8794392.
Drews, F.A., Pasupathi, M., Strayer, D.L., 2008. Passenger and cell phone conversations in simulated driving. J. Exp. Psychol. Appl. 14 (4), 392–400. https://doi.org/10.1037/a0013119.
Endsley, M.R., 2017. Autonomous driving systems: A preliminary naturalistic study of the Tesla Model S. J. Cogn. Eng. Decis. Mak. 11 (3), 225–238. https://doi.org/10.1177/1555343417695197.
Endsley, M.R., 1995. Measurement of situation awareness in dynamic systems. Hum. Factors 37 (1), 65–84. https://doi.org/10.1518/001872095779049499.
Friedman, L., 2020. Tesla vehicle deliveries and autopilot mileage statistics. WWW Document]. URL. https://hcai.mit.edu/tesla-autopilot-miles-and-vehicles/.
Gwet, K.L., 2014. Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters. Advanced Analytics. LLC.
Habibovic, A., Lundgren, V.M., Andersson, J., Klingegård, M., Lagström, T., Sirkka, A., Fagerlönn, J., Edgren, C., Fredriksson, R., Krupenia, S., Saluäär, D., Larsson, P., 2018. Communicating intent of automated vehicles to pedestrians. Front. Psychol. 9. https://doi.org/10.3389/fpsyg.2018.01336.
Hecker, S., Dai, D., Van Gool, L., 2018. Failure prediction for autonomous driving. IEEE Intelligent Vehicles Symposium, Proceedings 1792–1799. https://doi.org/10.1109/IVS.2018.8500495.
Horne, J.A., Reyner, L.A., 1995. Driver sleepiness. J. Sleep Res. 4, 23–29. https://doi.org/10.1111/j.1365-2869.1995.tb00222.x.
Hubmann, C., Schulz, J., Becker, M., Althoff, D., Stiller, C., 2018. Automated driving in uncertain environments: Planning with interaction and uncertain maneuver prediction. IEEE Trans. Intell. Veh. 3 (1), 5–17. https://doi.org/10.1109/tiv.2017.2788208.
Jackson, L., Chapman, P., Crundall, D., 2009. What happens next? Predicting other road users' behaviour as a function of driving experience and processing time. Ergonomics 52 (2), 154–164. https://doi.org/10.1080/00140130802030714.
Kircher, K., Ahlstrom, C., 2017. Minimum required attention: A human-centered approach to driver inattention. Hum. Factors 59 (3), 471–484. https://doi.org/10.1177/0018720816672756.
Kircher, K., Ahlström, C., 2018. Evaluation of methods for the assessment of attention while driving. Accid. Anal. Prev. 114, 40–47. https://doi.org/10.1016/j.aap.2017.03.013.
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. Hum. Factors. https://doi.org/10.1177/0018720819866946.
Kujala, T., Saariluoma, P., 2018. Cognitive mimetics for designing intelligent technologies. Adv. Human-Computer Interact. https://doi.org/10.1155/2018/9215863.
Lake, B.M., Ullman, T.D., Tenenbaum, J.B., Gershman, S.J., 2017. Building machines that learn and think like people. Behav. Brain Sci. 40. https://doi.org/10.1017/S0140525X16001837.
Lee, Y.M., Madigan, R., Garcia, J., Tomlinson, A., Solernou, A., Romano, R., Markkula, G., Merat, N., Uttley, J., 2019. Understanding the messages conveyed by automated vehicles. In: Proceedings - 11th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2019. doi:. pp. 134–143. https://doi.org/10.1145/3342197.3344546.
Li, Y., Dikmen, M., Hussein, T.G., Wang, Y., Burns, C., 2018. To cross or not to cross: Urgency-based external warning displays on autonomous vehicles to improve pedestrian crossing safety. Proceedings of the 10th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications 188–197. https://doi.org/10.1145/3239060.3239082.
Lv, C., Cao, D., Zhao, Y., Auger, D.J., Sullman, M., Wang, H., Dutka, L.M., Skrypchuk, L., Mouzakitis, A., 2018. Analysis of autopilot disengagements occurring during autonomous vehicle testing. IEEE/CAA J. Autom. Sin. 5 (1), 58–68. https://doi.org/10.1109/JAS.2017.7510745.
Mahadevan, K., Somanath, S., Sharlin, E., 2018. Communicating awareness and intent in autonomous vehicle-pedestrian interaction. Conference on Human Factors in Computing Systems - Proceedings. https://doi.org/10.1145/3173574.3174003.
Markkula, G., Madigan, R., Nathanael, D., Portouli, E., Lee, Y.M., Dietrich, A., Billington, J., Schieben, A., Merat, N., 2020. Defining interactions: A conceptual framework for understanding interactive behaviour in human and automated road traffic. Theor. Issues Ergon. Sci. https://doi.org/10.1080/1463922X.2020.1736686.
Markkula, G., Romano, R., Madigan, R., Fox, C.W., Giles, O.T., Merat, N., 2018. Models of human decision-making as tools for estimating and optimizing impacts of vehicle automation. Transp. Res. Rec. 2672 (37), 153–163. https://doi.org/10.1177/0361198118792131.
Matthews, M.L., Bryant, D.J., Webb, R.D.G., Harbluk, J.L., 2001. Model for situation awareness and driving: Application to analysis and research for intelligent transportation systems. Transportation Research Record. pp. 26–32. https://doi.org/10.3141/1779-04.
Mayring, P., 2000. Qualitative content analysis. Forum Qual. Soc. Res. 1, 2.
McGoogan, C., 2016. Elon Musk: Tesla's Autopilot is twice as safe as humans. Telegr.
Meghjani, M., Luo, Y., Ho, Q.H., Cai, P., Verma, S., Rus, D., Hsu, D., 2019. Context and intention aware planning for urban driving. IEEE International Conference on Intelligent Robots and Systems. https://doi.org/10.1109/IROS40897.2019.8967873.
Merat, N., Lee, Y.M., Markkula, G., Uttley, J., Camara, F., Fox, C., Dietrich, A., Weber, F., Schieben, A., 2019. How do we study pedestrian interaction with automated vehicles? Preliminary findings from the European interACT project. In: Meyer, G., Beiker, S. (Eds.), Road Vehicle Automation 6. Springer. https://doi.org/10.1007/978-3-030-22933-7_3.
Miles, M.B., Huberman, A.M., 1994. Qualitative data analysis: An expanded sourcebook. Sage, Thousand Oaks, CA.
Mirnig, N., Perterer, N., Stollnberger, G., Tscheligi, M., 2017. Three strategies for autonomous car-to-pedestrian communication: A survival guide. ACM/IEEE International Conference on Human-Robot Interaction 209–210. https://doi.org/10.1145/3029798.3038402.
Newell, A., Simon, H.A., 1972. Human Problem Solving. Prentice-Hall, Englewood Cliffs, NJ.
Pollatsek, A., Narayanaan, V., Pradhan, A., Fisher, D.L., 2006. Using eye movements to evaluate a PC-based risk awareness and perception training program on a driving simulator. Hum. Factors 48 (3), 447–464. https://doi.org/10.1518/001872006778606787.
Portouli, E., Nathanael, D., Amditis, A., Lee, Y.M., Merat, N., Uttley, J., Giles, O., Markkula, G., Dietrich, A., Schieben, A., Jenness, J., 2019. Methodologies to understand the road user needs when interacting with automated vehicles. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 35–45. https://doi.org/10.1007/978-3-030-22666-4_3.
Press, A., 2018. Report: Tesla on autopilot is safer driver than people. The Columbian.
Rasmussen, J., 1983. Skills, rules, and knowledge; Signals, signs, and symbols, and other distinctions in human performance models. IEEE Trans. Syst. Man Cybern. SMC-13 3, 257–266. https://doi.org/10.1109/TSMC.1983.6313160.
Rasmussen, J., 1982. Human errors. A taxonomy for describing human malfunction in industrial installations. J. Occup. Accid. 4 (2–4), 311–333. https://doi.org/10.1016/0376-6349(82)90041-4.
Rasouli, A., Tsotsos, J.K., 2019. Autonomous vehicles that interact with pedestrians: A survey of theory and practice. IEEE Trans. Intell. Transp. Syst. 1–19. https://doi.org/10.1109/tits.2019.2901817.
Saariluoma, P., 1995. Chess Players' Thinking: A Cognitive Psychological Approach. Psychology Press.
Saariluoma, P., Kujala, T., Karvonen, A., Ahonen, M., 2018. Cognitive mimetics - Main ideas. In: 2018 World Congress in Computer Science, Computer Engineering and Applied Computing, CSCE 2018 - Proceedings of the 2018 International Conference on Artificial Intelligence. ICAI 2018. pp. 202–206.
Schmidt, E.A., Schrauf, M., Simon, M., Fritzsche, M., Buchner, A., Kincses, W.E., 2009. Drivers' misjudgement of vigilance state during prolonged monotonous daytime

driving. Accid. Anal. Prev. 41 (5), 1087–1093. https://doi.org/10.1016/j.aap.2009.06.007.
Schneemann, F., Gohl, I., 2016. Analyzing driver-pedestrian interaction at crosswalks: A contribution to autonomous driving in urban environments. IEEE Intelligent Vehicles Symposium, Proceedings 38–43. https://doi.org/10.1109/IVS.2016.7535361.
Schwarting, W., Alonso-Mora, J., Rus, D., 2018. Planning and decision-making for autonomous vehicles. Annu. Rev. Control. Robot. Auton. Syst. https://doi.org/10.1146/annurev-control-060117-105157.
Schwarting, W., Pierson, A., Alonso-Mora, J., Karaman, S., Rus, D., 2019. Social behavior for autonomous vehicles. Proc. Natl. Acad. Sci. https://doi.org/10.1073/pnas.1820676116.
Society of Automotive Engineers, 2019. Levels of driving automation. [WWW Document]. URL. https://www.sae.org/news/2019/01/sae-updates-j3016-automated-driving-graphic.
Stahl, P., Donmez, B., Jamieson, G.A., 2016. Supporting anticipation in driving through attentional and interpretational in-vehicle displays. Accid. Anal. Prev. 91, 103–113. https://doi.org/10.1016/j.aap.2016.02.030.
Stahl, P., Donmez, B., Jamieson, G.A., 2013. Anticipatory driving competence - Motivation, definition & modeling. In: Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. AutomotiveUI 2013. https://doi.org/10.1145/2516540.2516579.
Tanida, K., Pöppel, E., 2006. A hierarchical model of operational anticipation windows in driving an automobile. Cogn. Process. 7 (4), 275–287. https://doi.org/10.1007/s10339-006-0152-9.
Teoh, E.R., Kidd, D.G., 2017. Rage against the machine? Google's self-driving cars versus human drivers. J. Safety Res. 63, 57–60. https://doi.org/10.1016/j.jsr.2017.08.008.
Thiffault, P., Bergeron, J., 2003. Monotony of road environment and driver fatigue: A simulator study. Accid. Anal. Prev. 35 (3), 381–391. https://doi.org/10.1016/S0001-4575(02)00014-3.
Ting, P.H., Hwang, J.R., Doong, J.L., Jeng, M.C., 2008. Driver fatigue and highway driving: A simulator study. Physiol. Behav. 94 (3), 448–453. https://doi.org/10.1016/j.physbeh.2008.02.015.
van Huysduynen, H.H., Terken, J., Eggen, B., 2018. Why disable the autopilot? Proceedings of the 10th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications 247–257.
Ventsislavova, P., Crundall, D., 2018. The hazard prediction test: A comparison of free-response and multiple-choice formats. Saf. Sci. 109, 246–255. https://doi.org/10.1016/j.ssci.2018.06.004.
Ventsislavova, P., Crundall, D., Baguley, T., Castro, C., Gugliotta, A., Garcia-Fernandez, P., Zhang, W., Ba, Y., Li, Q., 2019. A comparison of hazard perception and hazard prediction tests across China, Spain and the UK. Accid. Anal. Prev. 122, 268–286. https://doi.org/10.1016/j.aap.2018.10.010.
Vinkhuyzen, E., Cefkin, M., 2016. Developing socially acceptable autonomous vehicles. Ethnogr. Prax. Ind. Conf. Proc. 1, 522–534. https://doi.org/10.1111/1559-8918.2016.01108.
Ward, N.J., 2000. Automation of task processes: An example of intelligent transportation systems. Hum. Factors Ergon. Manuf. 10 (4), 395–408. https://doi.org/10.1002/1520-6564(200023)10:4<395::aid-hfm3>3.3.co;2-v.
Wickens, C.D., 2008. Multiple resources and mental workload. Hum. Factors 50 (3), 449–455. https://doi.org/10.1518/001872008X288394.
Wissing, C., Nattermann, T., Glander, K.H., Bertram, T., 2018. Trajectory prediction for safety critical maneuvers in automated highway driving. IEEE Trajectory prediction for safety critical maneuvers in automated highway driving Conference on Intelligent Transportation Systems, Proceedings, ITSC 4–7. https://doi.org/10.1109/ITSC.2018.8569296.
Zaidel, D.M., 1992. A modeling perspective on the culture of driving. Accid. Anal. Prev. https://doi.org/10.1016/0001-4575(92)90011-7.

II

REFINING DISTRACTION POTENTIAL TESTING GUIDELINES BY CONSIDERING DIFFERENCES IN GLANCING BEHAVIOR

by

Hilkka Grahn & Toni Taipalus, 2021

Transportation Research Part F: Psychology and Behavior, 79, 23–34

https://doi.org/10.1016/j.trf.2021.03.009

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

A R T I C L E I N F O

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

A B S T R A C T

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

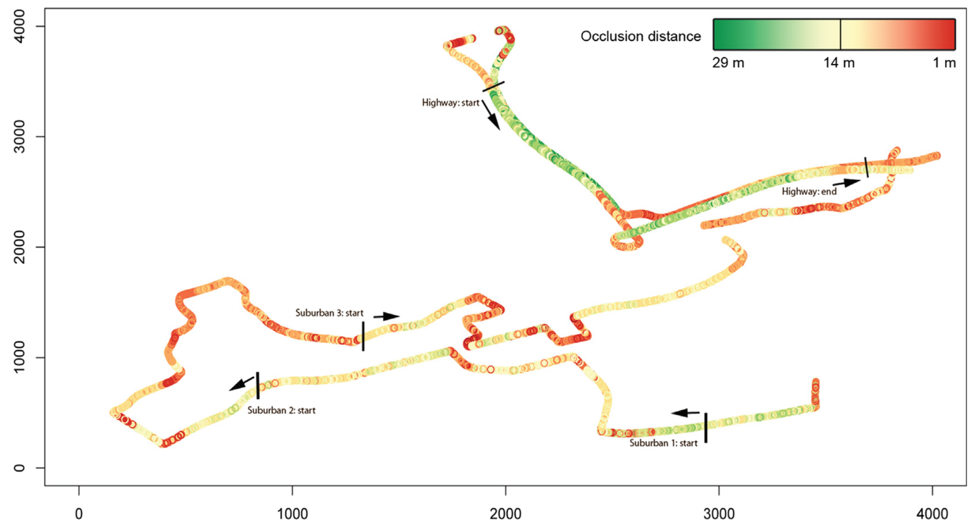

Fig. 1. Occlusion distance map (ODM) illustrates the routes used in the experiments – the color of a route point indicates the level of visual demand measured in occlusion distance (green = high occlusion distance, visually undemanding; red = low occlusion distance, visually demanding); the axes indicate 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

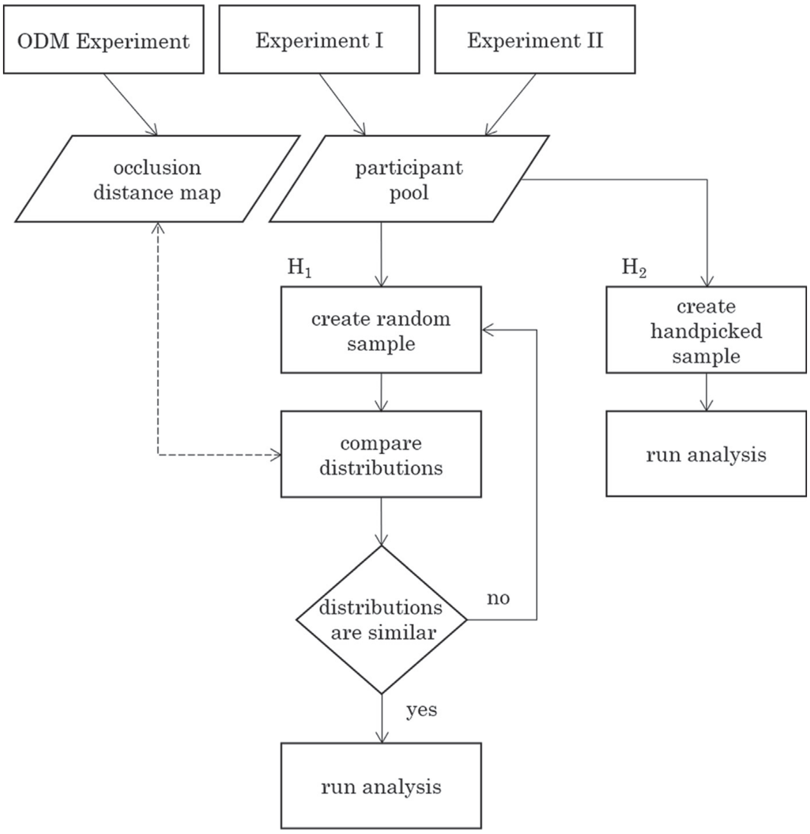

Fig. 2. The procedure – the previously conducted occlusion distance mapping (Kujala, Mäkelä et al., 2016) provides the occlusion distance map (ODM), and previously collected data for Experiments I and II (Grahn & Kujala, 2020) form the participant pool from which the samples for this study are selected – 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).

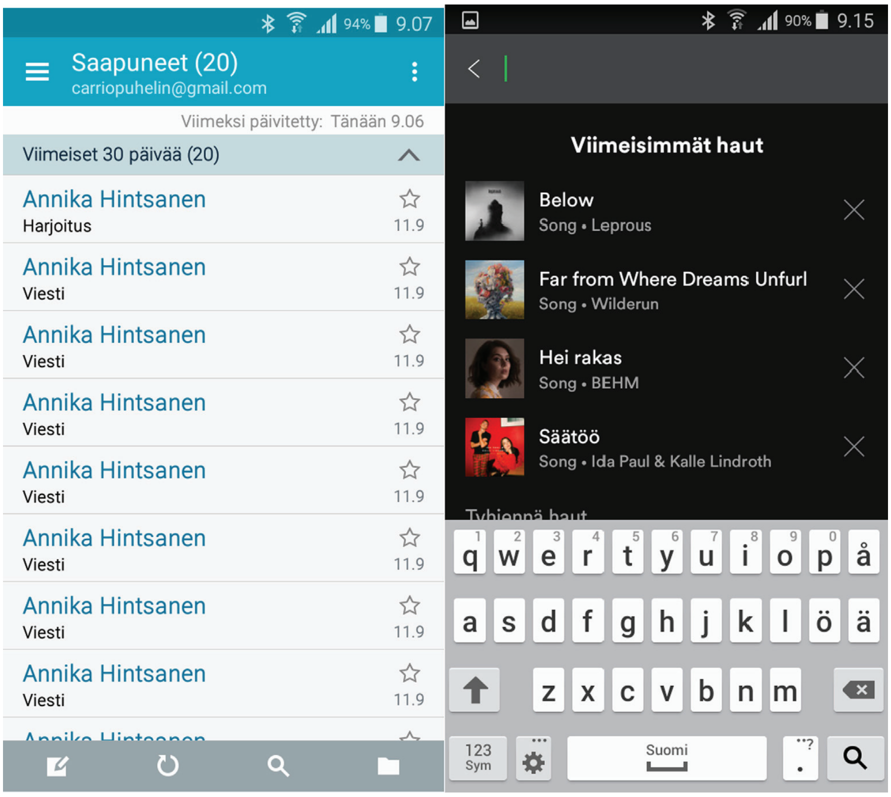

Fig. 4. Email reading and song searching in Experiment I.

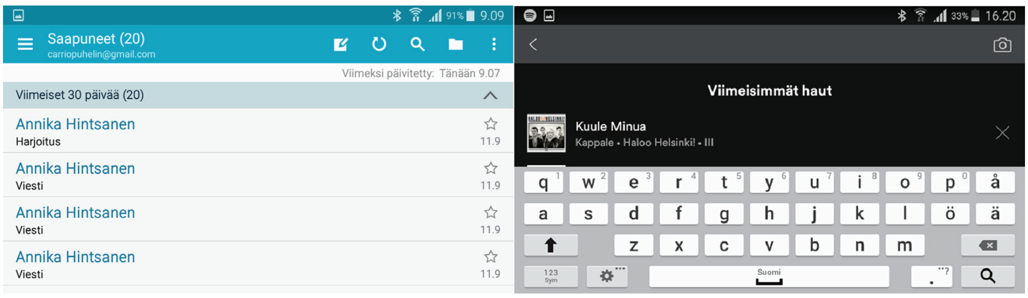

Fig. 5. Email reading and song searching in Experiment II (Grahn and Kujala, 2020).

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™: 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.

III

ON THE VISUAL DISTRACTION EFFECTS OF AUDIO-VISUAL ROUTE GUIDANCE

by

Tuomo Kujala, Hilkka Grahn, Jakke Mäkelä & Annegret Lasch, 2016

In Proceedings of the 8th International Conference on Automotive User Interfaces and Vehicular Applications, 169–176

https://doi.org/10.1145/3003715.3005421

Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '16), October 24–26, 2016, Ann Arbor, MI, USA

On the Visual Distraction Effects of Audio-Visual Route Guidance

Tuomo Kujala
University of Jyväskylä
Jyväskylä, Finland
tuomo.kujala@jyu.fi

Hilkka Grahn
University of Jyväskylä
Jyväskylä, Finland
hilkka.grahn@jyu.fi

Jakke Mäkelä
University of Jyväskylä
Jyväskylä, Finland
jakke.makela@gmail.com

Annegret Lasch
Google
Munich, Germany
annegret.lasch@gmail.com

ABSTRACT
This is the first controlled quantitative analysis on the visual distraction effects of audio-visual route guidance in simulated, but ecologically realistic driving scenarios with dynamic maneuvers and self-controlled speed ($N = 24$). The audio-visual route guidance system under testing passed the set verification criteria, which was based on drivers' preferred occlusion distances on the test routes. There were no significant effects of an upcoming maneuver instruction location (up, down) on the in-car display on any metric or on the experienced workload. The drivers' median occlusion distances correlated significantly with median in-car glance distances. There was no correlation between drivers' median occlusion distance and intolerance of uncertainty but significant inverse correlations between occlusion distances and age as well as driving experience were found. The findings suggest that the visual distraction effects of audio-visual route guidance are low and provide general support for the proposed testing method.

Author Keywords
Driver distraction; navigation system; visual demand; visual occlusion; occlusion distance; intolerance of uncertainty.

ACM Classification Keywords
H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces – Graphical user interfaces (GUI), Evaluation/methodology

INTRODUCTION
Widely used verification guidelines for visual-manual in-vehicle electronic devices of the National Highway Traffic Safety Administration (NHTSA) [9] recommend three metrics to verify the distraction potential caused by electronic devices (total and mean duration of in-car glances, percentage of over-2-second in-car glances). According to the guidelines, the testing scenario should consist of a straight highway driven with steady speed and keeping of a static distance (70 m) to a lead car.

Automotive'UI 16, October 24-26, 2016, Ann Arbor, MI, USA
ACM 978-1-4503-4533-0/16/10.
http://dx.doi.org/10.1145/3003715.3005421

However, real-world driving scenarios with route guidance include turns, lane selections, and the associated decelerating and accelerating behaviors, and thus, the distraction effects of navigation system cannot be reliably tested within the NHTSA scenario [6]. In real-world driving, visual route guidance is in fact realized typically in these types of dynamic situations when approaching turns or selecting lanes. Furthermore, the safety risk of a 2-second in-car glance is highly dependent on the driving speed among other situational variables. The visual distraction effects of a 2-second off-road glance are significantly different while driving 60 than 80 kilometers per hour in a particular road environment [7].

The NHTSA verification guidelines [9] have received a lot of attention related to the effects of participant sampling on the outcomes of the testing (pass or fail, see e.g., [1,3]). Validation of relevant driver sample characteristics seem to be missing, and it has been argued that it depends mostly on the random driver sample if an in-car task passes or fails the criteria, that is, if the drivers happen to be 'short-glancers' or 'long-glancers'.

For the reasons above, the present study follows an alternative test environment and in-car glance duration verification criteria for dynamic and self-paced driving scenarios suggested by Kujala and Mäkelä (2015, [6]). This is the first time the proposed method is applied to a real in-car task. The verification criteria are based on visual occlusion data mapped on the test routes.

The technique of visual occlusion is an established method to define the visual demands of driving [10,13]. In the technique, the visual field of the driver is intermittently occluded with an occlusion visor or opaque screens. The occlusion time (OT) and/or the occlusion distance (OD) the driver is able to drive without visual information of the forward roadway are measured. By the means of visual occlusion, the testing method presented in [6] utilizes the preferred occlusion distances of 97 drivers on simulated real Finnish roads [7] as a baseline (i.e., control) of acceptable in-car glances. Occlusion distance (OD) refers to the distance that a driver feels comfortable to drive with occluded vision while fully concentrating on the driving task [7]. Basically, the lower the occlusion distance for a road, the higher the environmental visual demands of the road.

In the testing method of Kujala and Mäkelä [6], it is possible to define threshold ODs, which drivers are not willing to exceed while fully concentrating on the driving task, for each point of a route used for testing. Exceeding the threshold OD when looking at an in-car device indicates a failure in control of the visual sampling off road. In this situation, the driver is driving a longer distance without focal visual information of the road than the driver would be willing to drive occluded when fully focusing on the driving task.

A traffic light analogy is used by Kujala and Mäkelä [6] to define the acceptability of the individual in-car glances from the sample ODs. The specific threshold values are:

- Green: the median OD of the driver sample ODs for a given road point gives a value that is assumed to be still safe (drivers are assumed to have behaved cautiously on average).
- Red: the 85th percentile OD gives a value that when exceeded, the glance can definitely be considered as unacceptable by the majority of the driver population.

A clear goal for an acceptable in-car task is to have all the glances within the green category but some tolerance must be accepted, as there are clear individual differences in the preferred occlusion distances [7]. Unfortunately, the authors [6] do not provide explicit verification criteria for acceptable percentages per glance category. For this reason, we took their original data (N = 97) and calculated what are the median percentages of occlusion distances below or at the median OD ('green occlusions') and occlusion distances exceeding the 85th percentile OD ('red occlusions') of the drivers. Based on these percentages, the verification threshold of green glances was set to 68%. That is, at least 68% of in-car glances should be shorter or at most at the median ODs (i.e., green). Accordingly, the verification threshold of red glances was set to 6%. That is, at most 6% of in-car glances can be above the 85th percentile ODs (i.e., red). In order for these thresholds to be representative, it is vital that the preferred occlusion distance distribution of a test sample is comparable to the original sample (N = 97), from which these threshold values are derived from.

The purpose of the current study was to run the first controlled quantitative analysis on the visual distraction effects of audio-visual route guidance in realistic driving scenarios with dynamic turns and self-controlled speed, and to further validate the test method of Kujala and Mäkelä [6]. The study is divided into two parts: distraction testing and occlusion experiment. The behaviors of the same participants were studied in the both parts of the study.

In the distraction testing we studied a commercial audio-visual route guidance prototype in a motion-platform driving simulator following the testing and verification criteria described in [6]. We investigated the visual distraction potential between two alternative navigation system display designs; maneuver box location up or down. Maneuver box contains information about the next maneuver and the distance to the turn (see Figure 1).

The research questions for the distraction testing included:

1. Does the studied commercial prototype for route guidance pass the set verification criteria?
2. Are there significant differences in the visual distraction potential between the two alternative display designs (maneuver box up or down)?
3. Do the drivers experience more task workload with either of the two designs and how does the workload compare to the occlusion experiment (NASA-TLX; [4])?
4. Are the test results comparable on routes with different visual demands (suburban and highway)?

Based on previous research (for review, see [12]) and because crash statistics haven't shown significant effects of similar route finding tasks on crash risk, we hypothesized that the task to follow the audio-visual route guidance would pass the verification criteria regardless of the visual user interface design.

From a design point of view, the advantage of the maneuver box's lower position is that it allows to display also the following maneuvers in a natural position above the first one (Figure 1). However, we expected that the upper location of the maneuver box enables the driver to sample more efficiently the upcoming maneuver guidance and upcoming route on the map with a single glance, whereas larger movements of gaze are expected when the maneuver box is located farther apart down from the upcoming route. This could lead to longer individual in-car glance durations with the maneuver box down, if these two types of visual information are sampled within a glance.

For the method validation, the second part of the study was an occlusion experiment following [7]. The research questions were:

1. Is the OD distribution of our test sample comparable to the OD distribution for highway driving reported in [7] (the baseline data)?
2. Do the ODs correlate with the driver's preferred in-car glance distances (i.e., distance traveled during an in-car glance) across the in-car tasks?
3. Do the drivers' ODs correlate with their reported intolerance of uncertainty [2], age or driving experience? Intolerance of uncertainty is one plausible personality factor that could explain at least partly the individual differences in the preferred occlusion distances.

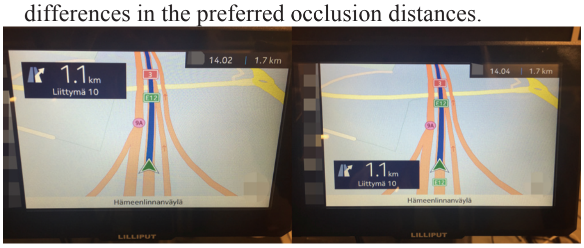

Figure 1: Maneuver boxes on the in-car display (left up, right down).

Figure 2: The experimental setup.

METHOD

The experimental design of the distraction test was within-subjects 2 x 2. The independent variables were the maneuver box location (up, down, Figure 1) and the driving environment (highway, suburban). In the occlusion experiment the independent variables included age, driving experience, and the intolerance of uncertainty [2].

Participants

In total we had 24 participants who were recruited via e-mail lists. We tried to follow the NHTSA [9] recommendations on the driver sample as closely as possible. Twelve of the participants were male and twelve were female. Participants' age varied from 21 to 67 years and mean age was 38,4 (*SD* = 15.3). Six of the participants were 18 to 24 years old, six 25 to 39 years old, eight 40 to 55 years old, and four over 55 years old. Half of the participants in each age group were male and half female. The age categories followed the NHTSA test participant recommendations [9]. The reason of the small deviation is because two older (55+) participants suffered from minor simulator sickness and were replaced with over 40 years old participants.

All participants had a valid driver's license and they drove at least 5,000 kilometers per year. The total distance driven varied from 5,000 to 30,000 kilometers (*M* = 13,300, *SD* = 6,900). The lifetime driving experience of the participants varied from four years to 49 years (*M* = 19.5, *SD* = 14.5). All participants had normal or corrected-to-normal vision and were able to drive and navigate without glasses. Experiments were instructed in Finnish and all participants were fluent Finnish-speakers. All participants were unfamiliar with the prototype navigation system that was tested. Each participant was rewarded with a movie ticket and a car phone holder.

Apparatus

The experiments took place at the driving simulator laboratory at the University of Jyväskylä. The simulator can be categorized as medium-fidelity and it consisted of CKAS Mechatronics 2-DOF motion platform, longitudinally adjustable seat as well as Logitech G27 force-feedback steering wheel and pedals (Figure 2). Automatic transmission was used during the experiments. The driving scene was displayed on three 40" LED screens (95.6 cm x 57.4 cm) and the resolution was 1440 x 900 pixels per screen. The middle screen included head-up display speedometer, RPM gauge and a rear-view mirror. Both side screens had side mirrors. For the occlusion experiment, the back of the steering wheel was equipped with a lever for removing the occlusion of the driving scene for 500 milliseconds for each press during the visual occlusion trial. Continuous pressing of the lever kept the driving scene

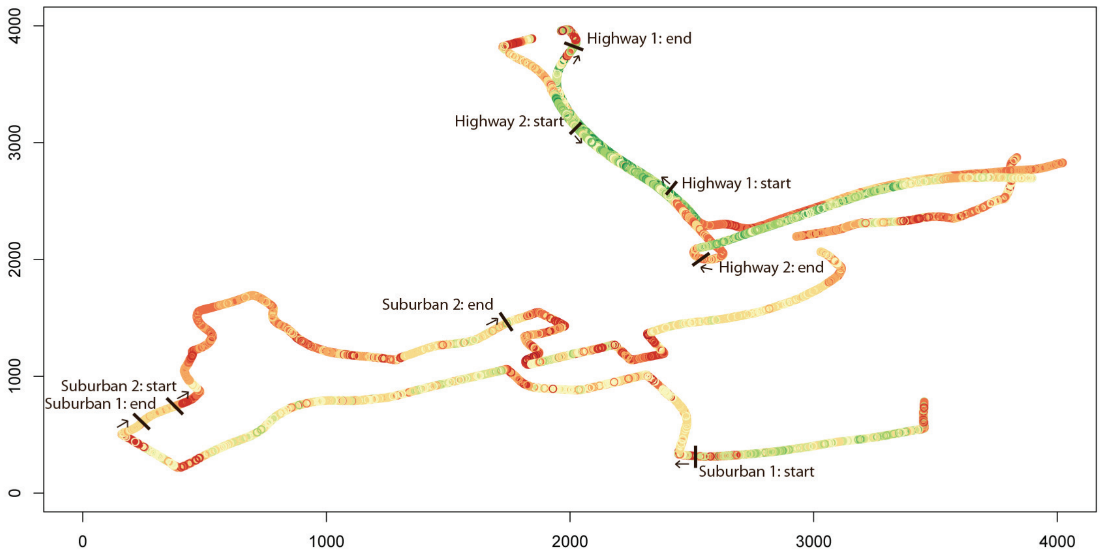

Figure 3: The pre-defined routes for the experiments. Green: low visual demands (high OD), red: high visual demands (low OD). NB. Roads are not in scale.

continuously visible. The driving simulation software was provided by Eepsoft (http://www.eepsoft.fi/). The software saved driving log data at 10 Hz. The used predefined routes simulated real Finnish highway and suburban roads located at Martinlaakso, Vantaa. The same roads were used as in the study of Kujala and Mäkelä [6] (see Figure 3). The driving simulator sent real-time simulated GPS data to the navigation system under testing to support route guidance.

The prototype navigation system was running on Intel NUC5i3RYK and displayed on Lilliput 779GL-70NP/C/T - 7" capacitive touchscreen display. Ergoneers' Dikablis 50 Hz eye-tracking system was used to record participants' eye movements. LAN bridge was used for the synchronization of the driving simulator and eye-tracking data.

Procedure

After signing an informed consent, participants were taken to the driving simulator and the seat was adjusted for each participant. At first, participants practiced driving in an artificial city environment with other road users as long as they wanted. The average practice time was 3.2 minutes. After they felt comfortable with driving, they did a second practice on a suburban route in Martinlaakso with audio-visual route guidance on to get familiar with the guidance that the navigation system provided (two turns). When both practices were done, the eye-tracking headset was put on, adjusted and calibrated.

During the distraction testing, participants were instructed to follow the audio-visual route guidance to find the pre-defined destination. The participants were able to listen the route guidance as well as to see the route and the upcoming maneuvers displayed on a map on the touchscreen display. The routes were set by the experimenter, thus there was no manual input required from the driver during the trials. The participants were told to try to drive about 80 kilometers per hour on highway and about 50 kilometers per hour on suburban roads (speed limits), but that they can control the speed freely according to the situational demands. The routes were always driven in the same order, but counter-balancing was done for the design alternatives. In order to control the possible learning effects, there were two slightly different highway (one turn each; a ramp) and two slightly different suburban routes (five turns each), see Figure 3. The routes were selected by finding routes as similar as possible regarding their visual demands (for both route types). The order of the routes was counter-balanced across the sample. In total each completed four trials: two on a highway scenario (1.4 and 1.5 km) and two on a suburban scenario (3.7 and 2.5 km). There were no other road users in the scenarios. After each trial, NASA-TLX questionnaire [4] was filled out.

Fourteen out of 24 participants (58.3%) got lost at least on one route. Most of the cases happened when the participant took a wrong turn just before the correct one. A possible reason for the high percentage could be the difficulty to assess distances in a driving simulator. The trials were rerun, unless the participant got lost in the last turn.

After the distraction testing, the visual occlusion experiment started. At first, the participants practiced in a city environment with other road users how to drive with vision occasionally occluded and how to use the lever that removed the occlusion of the driving scene. After the practice, the actual occlusion trial started. The average practice time was 3.3 minutes.

In the occlusion experiment, the screens were blank as default. By pressing the lever on the back of the steering wheel, the participant could remove the occlusion of the driving scene for 500 milliseconds at a time, following the original method by Senders et al. [10]. The findings of Senders et al. as well as Kujala et al. [7] suggest that 500 milliseconds of intermittent visibility is enough for at least experienced drivers to drive fairly fluently and according to traffic regulations in the studied scenarios.

The participants were instructed to follow the traffic regulations and drive safely but within these limits, try to drive vision occluded as long as possible. They were also told that six test participants who drive most accurately and the longest periods with occluded vision, get a second movie ticket. This was done in order to make the participants focus on the driving task but still try to maximize the preferred occlusion distance. The visual occlusion trial included highway routes only without other road users. The same highway routes were used as in the distraction test. The speed limits changed during the trial (from 80 to 120 km/h) and each limit was told to each participant by the experimenter at the same point of the route. Finally, the participants filled out the NASA-TLX questionnaire for the occlusion trial and the intolerance of uncertainty questionnaire [2].

Analysis

In the distraction testing, dependent variables were:

- Percentage of green in-car glances: in-car glance distances below or at median OD (for any 1x1-meter road point).
- Percentage of red in-car glances: in-car glance distances above 85th percentile OD (for any 1x1-meter road point).
- Total and mean duration of in-car glances, as well as the percentage of over-2-second in-car glances (after NHTSA [9] verification criteria, for comparison).
- Median in-car glance distance, that refers to distance traveled during an in-car glance.
- Reduced NASA-TLX (no weighting) for measuring experienced task workload (for each condition).

In-car glance durations were scored in real-time by a script reading the pupil's x and y coordinates from the eye-tracker, and logged with the location data provided by the driving simulator. The durations were scored following the SAE-J2396 definition [11] with the addition of gaze transition time back to driving scene, in order to enable more direct comparability with OD. Each glance was manually inspected from a synchronized video (25 fps) by a data reducer for validity using Noldus Observer XT software. All the trials with inaccuracies were manually scored frame-by-frame from the video material. Perfect automated glance recognition was in 33 out of 96 trials (34.4 %). In total 38 out of 96 trials were manually scored (39.6 %). The automated glance scoring made some false positive in-car glances (mainly glances on the side mirror and the speedometer), but those were manually removed from the data (in 26.0 % of trials). Blinks were removed from the data by rejecting glances shorter than 300 ms.

One-sample sign test was used to test the equality of median green and red in-car glance durations on the two design alternatives (maneuver box locations) to the set verification thresholds (68% and 6%). The differences between the maneuver box locations in the percentages of green and red glances as well as in over-2-second in-car glances were also tested with the sign test due to non-normal and asymmetric distributions. In addition, paired samples t-test was used for analyzing differences in mean in-car glance durations and total in-car glance durations between the two maneuver box design alternatives. One-way repeated measures ANOVA was used to test differences in the experienced task workload between the route guidance trials (maneuver box up, down) and the occlusion trial (highway only). Greenhouse-Geisser correction was applied when the sphericity assumption was violated. Bonferroni correction was applied for pairwise comparisons. Partial eta-squared and Cohen's d are reported as metrics of effect size where applicable.

In the occlusion experiment, dependent variables were median occlusion time (OT) and distance (OD). Median was used instead of mean because of the non-gaussian distributions of the occlusion metrics. Only occlusions made on the highway when the speed was over 20 m/s (72 km/h) were included in order to control for the effects of accelerations and decelerations in the start, junctions, and the end of the trial. Pearson product-moment correlation coefficient was used to test the correlation between median occlusion distance and median in-car glance distance, as well as the correlation between median occlusion distance and intolerance of uncertainty, driving experience, and age.

RESULTS

Distraction testing – Verification

Due to low number of in-car glances on the highway routes (M = 8, SD = 5, only one turn per route), it is more reliable to use the suburban routes (up: M = 34, down: M = 39, five turns per route) for the verification testing. Medians for the percentage of green glances were up: 67.9% and down: 60.5% (Figure 4). The verification threshold of green glances was set to 68%. One-sample sign test indicated no significant differences from 68% ("median equals to 68%", N = 24) for either condition: p > .999 (up) and p = .307 (down).

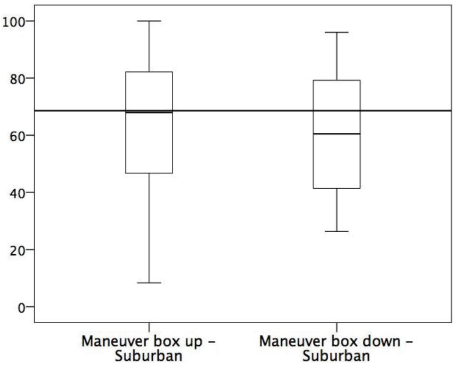

Figure 4: The percentage of green in-car glances. Verification threshold illustrated at 68% (median should be at or above).

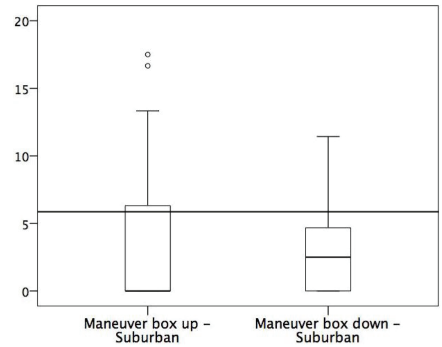

Figure 5: The percentage of red in-car glances. Verification threshold illustrated at 6% (median should be at or below).

The percentages of red glances were very low in general. Medians for the percentage of red glances were 0.0% for the up-condition and 2.5% for the down-condition (Figure 5). The verification threshold was set to 6%. One-sample sign test indicated the percentage of red glances was significantly lower from 6% for the maneuver box down condition, p < .001 ("median equals to 6%", N = 24). For the maneuver box up condition the difference from 6% was not significant, p = .152. However, note that also the up-condition passed the test, as the median percentage was not significantly higher than 6% (median for up was 0.0%).

Maneuver box location

We did not find significant effects of the maneuver box location on the metrics of [6], see Figures 4 and 5 (green glances: p > .999, red glances: p = .383, N = 24). For comparison, we also wanted to see if there were significant differences between the two alternative designs with the metrics of NHTSA [9] (for the suburban routes). However, note that the NHTSA metrics are not directly applicable here due to the dynamic and self-paced driving scenarios.

No significant effects of the maneuver box location on total in-car glance durations were found (p = .153, N = 24). Notable are the long total in-car glance durations for the tasks (up: M = 27.6 s, SD = 11.3; down: M = 31.6 s, SD = 17.2), well exceeding the NHTSA [9] recommendation of

12 s (max). However, this metric is directly dependent on the experimental design; on how many turns there are to make until reaching the destination, and not applicable here. Mean in-car glance durations stayed well below 1 second (up: M = .82 s, SD = .12; down: M = .82 s, SD = .16). There was no significant effect of the maneuver box location on mean in-car glance durations (p = .839, N = 24). The percentage of over-2-second in-car glances was very low (up: M = 0.35%, SD = .96; down: M = 0.29%, SD = 1.41), further indicating low visual demands of the route following tasks. There was no significant effect of the maneuver box location on the percentage of over-2-second in-car glances (p > .999, N = 24).

Experienced workload: NASA-TLX

There were no significant effects of the maneuver box location on the experienced workload on highway driving (Figure 6, mean difference down-up: .07, p = .980). However, the occlusion trial was experienced significantly more demanding than the route guidance trials ($F(1.57,36.18) = 53.70$, $p < .001$, partial η^2= .700). There were significant differences on the experienced workload between the occlusion trial and the maneuver box up trial (mean difference: 31.11, p < .001), and between the occlusion trial and the maneuver box down trial (mean difference: 31.04, p < .001). The experienced workload was at a highly similar level across the highway and suburban routes in the distraction testing (up: p = .697, down: p = .831).

Comparability across routes with different visual demands

When comparing the suburban test results with the highway test results, we found the same insignificant relative effects of the maneuver box location, similar mean in-car glance durations (up: .82 s, down: .85 s) and similar very low percentages of over-2-second in-car glances (~0%). Yet the percentages of red glances were higher (median up: 4.5%, median down: 15.5%) and the percentages of green glances were lower (median up: 38.8%, median down: 41.7%) than on the suburban routes. However, the highway data can be considered as unreliable due to low number of in-car glances (M = 8 glances for both conditions).

Occlusion experiment – Occlusion times and distances

The distributions of the drivers' median occlusion times (OT, Figure 7) and distances (OD, Figure 8) on the highway were fairly similar with the distributions reported in [7], with a slight skew towards the lower ODs. Median OD ranged from 6.2 to 50.4 m.

Occlusion distance vs. in-car glance distance

In-car glance distance refers to the distance that is traveled during an in-car glance. Median in-car glance distances correlated significantly between the suburban up and down trials (r = .633, p = .001) and were averaged for data reduction and comparison with median ODs. We found significant correlation between the drivers' median OD and median in-car glance distance, r = .47, p = .020 (Figure 9).

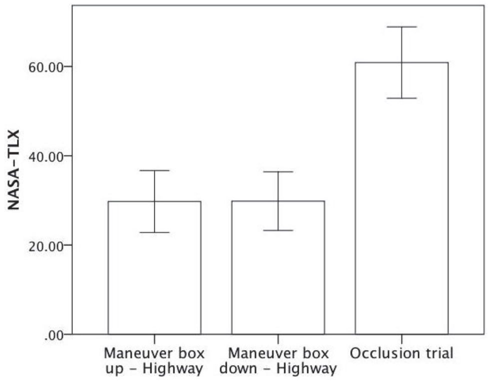

Figure 6: Total workload – highway, NASA-TLX (max. 100). Error bars: 95% CI.

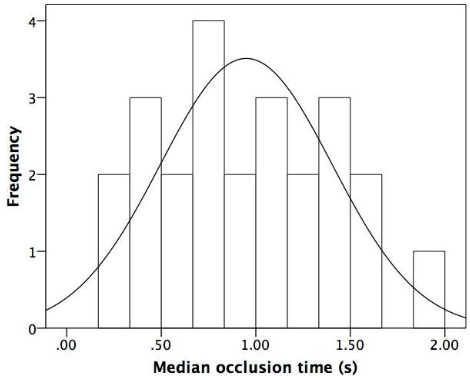

Figure 7: Median occlusion times (s) on highway (speed > 72 km/h).

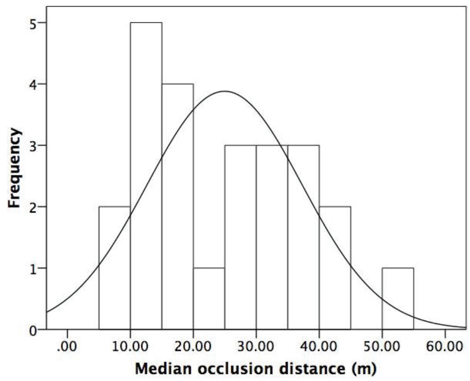

Figure 8: Median occlusion distances (m) on highway (speed > 72 km/h).

Occlusion distance vs. intolerance of uncertainty

We found no correlation between median occlusion distances and intolerance of uncertainty (r = .034, p = .873). However, there were significant inverse correlations between age and median OD (r = -.653, p = .001, Figure 10) as well as between driving experience and OD (r = -.637, p = .001). Here, driving experience correlated strongly with age (r = .993, p < .001).

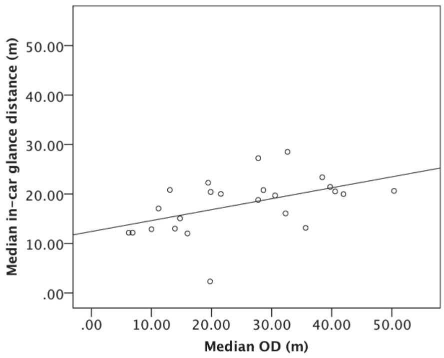

Figure 9: Median in-car glance distance (m) on the suburban routes vs. median OD (m) on highway.

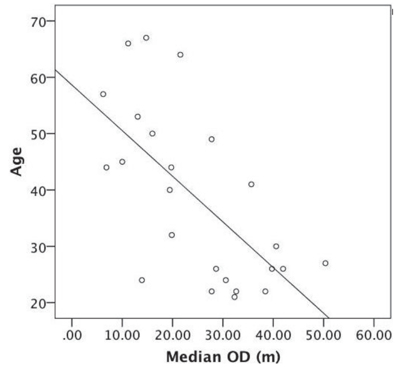

Figure 10: Age vs. median OD (m, highway).

GENERAL DISCUSSION

We studied the visual distraction effects of audio-visual route guidance in ecologically realistic driving scenarios with dynamic maneuvers and self-controlled speed. In the distraction testing part of the study we studied a commercial audio-visual route guidance system prototype following the testing and verification criteria described in [6]. The testing method enabled tactical control of multitasking for the participants, but in controlled settings.

Based on the percentages of red and green glances, the route guidance user interfaces under testing passed the set verification criteria. The audio-visual route guidance aids passed the test even if glances on the in-car display were required mostly on the visually high-demanding parts of the routes (i.e., before and at the turns). The current data can act as a baseline for an acceptable in-car task to which more complex in-car tasks, involving also visual-manual interactions, can be benchmarked against.

There were no significant effects of the maneuver box location (up, down) on the in-car display on any metric. In addition, there were no significant differences on the experienced workload between the two different maneuver box locations. The possible effects of showing more upcoming maneuvers above the maneuver box (down) should be studied.

We found differences in the verification metrics between the suburban and highway scenarios, but the highway test results can be considered unreliable due to low number of in-car glances (highway routes involved only one turn per route vs. five turns on the suburban routes). This finding suggests that a representative number of in-car glances should be collected per in-car task in order to get reliable verification data. For short tasks, the tasks should be repeated so that at least 20 in-car glances can be examined for red and green glances. In this way, single red in-car glances do not distort the percentages. There are definitely more complex road environments to navigate than the ones in the study but many of the participants took wrong turns (trials excluded in the results), which could suggest the route finding task itself was fairly difficult. However, the effects of the test route on the verification test results should be more carefully studied in following research.

The second part of the study was focused on method validation by the means of visual occlusion. The distributions of the drivers' median occlusion times and distances on the highway were fairly similar with the distributions reported in [6], and thus suggesting a representative sample of the driver population. This kind of sample validation must be an integral part of the verification testing, and is missing from, for instance, in the current NHTSA [9] test guidelines. It seems the NHTSA verification results are highly dependent on the distribution of 'short-glancers' and 'long-glancers' in the sample [1]. This finding is understandable given the large variance in the preferred occlusion times and distances the drivers are willing to tolerate (see Figures 9 and 10, see also [7] and [8]).

We found a significant correlation between the drivers' median occlusion distance and median in-car glance distance. This finding gives support for the testing method [6]; the idea that the drivers' self-preferred occlusion distances can be used as a comparison point for appropriate visual in-car glancing behavior - and thus, visual distraction. However, this finding should be replicated with other types of in-car tasks. Moreover, the correlation tells us there are differences in either drivers' uncertainty (or risk) tolerance levels, or that other drivers are just more apt to drive longer distances essentially blind than others. In the latter case, it would be fairer to compare each driver's in-car glance distances to his or her individual preferred ODs. However, like discussed in [6], the high-OD drivers in [7] were associated with decreased lane-keeping performance, suggesting overestimation of their visual sampling skills.

The occlusion experiment represents a baseline comparison point for a driving scenario with a 'maximum level of tolerated visual inattention' while focusing on driving only. NASA-TLX for highway driving indicated that the occlusion trial was experienced significantly more demanding than the route guidance trials. This is in line with the test data, suggesting the demands of the route guidance following were at a low level. When it comes to the intolerance of uncertainty [2], we found no correlation between drivers' median occlusion distances and intolerance of uncertainty. This suggests that the general intolerance of uncertainty - a personality trait - is not one of the factors behind the individual differences on the preferred occlusion distances or on the in-car glance distances. Instead, both age as well as driving experience had significant inverse correlations with median occlusion distance. These are related variables, and the exact factors behind the preferred occlusion distances are a topic for further research. These could be, for instance, the spatial span of working memory [5], that has been observed to decrease with age, or some skill acquired with increasing driving experience. However, the study of [7] with a larger sample did not indicate significant effects of driving experience on OD. In the study, neither did age correlate significantly with OD (p = .090), but the sample was not equally distributed across different age groups. However, in order to have a comparable distribution of 'short-glancers' and 'long-glancers' in a test group, the NHTSA [9] guidelines on the age distribution of the drivers can be recommended.

Finally, it should be noted that the testing method applies only to driving scenarios with empty roads [6]. The road environment-based ODs are not a reliable baseline if there is other traffic on the roads, as the traffic will likely increase the visual demands of driving. Future research should address how one could define baseline ODs (or OTs) for even more dynamic traffic scenarios including other traffic.

CONCLUSION

This was the first controlled quantitative analysis on the visual distraction effects of audio-visual route guidance in simulated, but ecologically realistic driving scenarios with dynamic maneuvers and self-controlled speed. The results suggest that the visual distraction effects of audio-visual route guidance are low. The findings provide general support for the testing method, which uses drivers' preferred occlusion distances on the selected test routes as the baseline for acceptable in-car glance durations.

ACKNOWLEDGMENTS

The study was funded by HERE Global B.V.

REFERENCES

1. Mikael L. Aust, Annie Rydström, Robert Broström, and Trent Victor. 2015. Better ways to calculate pass/fail criteria for the eye glance measurement using driving simulator test. In *Proc. of the 4th International Conference on Driver Distraction and Inattention.*
2. Jane Birrell, Kevin Meares, Andrew Wilkinson and Mark Freeston. 2011. Toward a definition of intolerance of uncertainty: A review of factor analytical studies of the Intolerance of Uncertainty Scale. *Clinical Psychology Review*, 31, 7, 1198-1208. http://dx.doi.org/10.1016/j.cpr.2011.07.009
3. Robert Broström, Mikael L. Aust, Linnea Wahlberg and Laban Källgren. 2013. What drives off-road glance durations during multitasking–capacity, practice or strategy. In *Proc. of the 3rd International Conference on Driver Distraction and Inattention,* 4-6.
4. Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX: Results of empirical and theoretical research. In *Human Mental Workload,* Peter A. Hancock and Najmedin Ed Meshkati (eds.). Elsevier, Amsterdam, 139-183.
5. Lisa Jenkins, Joel Myerson, Jennifer A. Joerding and Sandra Hale. 2000. Converging evidence that visuospatial cognition is more age-sensitive than verbal cognition. *Psychology and Aging*. 15, 1, 157-175. http://dx.doi.org/10.1037/0882-7974.15.1.157
6. Tuomo Kujala and Jakke Mäkelä. 2015. Development of a testing environment and a verification procedure for in-car tasks with dynamic driving scenarios. In *Proc. of the 4th International Conference on Driver Distraction and Inattention*.
7. Tuomo Kujala, Jakke Mäkelä, Ilkka Kotilainen and Timo Tokkonen. 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. http://dx.doi.org/10.1177/0018720815595901
8. Ja Young Lee, Madeleine Gibson and John D. Lee 2015. Secondary task boundaries influence drivers' glance durations. In *Proc. of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 273-280. http://dx.doi.org/10.1145/2799250.2799269
9. National Highway Traffic Safety Administration. 2013. *Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices* (NHTSA-2010-0053). Washington, DC: NHTSA.
10. John W. Senders, A. B. Kristofferson, W. H. Levison, C. W. Dietrich and J. L. Ward. 1967. The attentional demand of automobile driving. *Highway Research Record*, 195.
11. 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.
12. Louis Tijerina, S. Johnston, E. Parmer, M. D. Winterbottom and M. Goodman. 2000. *Driver distraction with wireless telecommunications and route guidance systems* (No. HS-809 069). Washington, DC: NHTSA.
13. Omer Tsimhoni and Paul Green. 1999. Visual demand of driving curves determined by visual occlusion. In *Proc. of Vision in Vehicles 8 Conference*.

IV

VISUAL DISTRACTION EFFECTS OF IN-CAR TEXT ENTRY METHODS: COMPARING KEYBOARD, HANDWRITING AND VOICE RECOGNITION

by

Tuomo Kujala & Hilkka Grahn, 2017

In Proceedings of the 9th International Conference on Automotive User Interfaces and Vehicular Applications, 1–10

https://doi.org/10.1145/3122986.3122987

Proceedings of the 9th ACM International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '17), September 24–27, 2016, Oldenburg, Germany .

Visual Distraction Effects of In-Car Text Entry Methods – Comparing Keyboard, Handwriting and Voice Recognition

Tuomo Kujala
University of Jyväskylä
Jyväskylä, Finland
tuomo.kujala@jyu.fi

Hilkka Grahn
University of Jyväskylä
Jyväskylä, Finland
hilkka.grahn@jyu.fi

ABSTRACT
Three text entry methods were compared in a driving simulator study with 17 participants. Ninety-seven drivers' occlusion distance (OD) data mapped on the test routes was used as a baseline to evaluate the methods' visual distraction potential. Only the voice recognition-based text entry tasks passed the set verification criteria. Handwriting tasks were experienced as the most demanding and the voice recognition tasks as the least demanding. An individual in-car glance length preference was found, but against expectations, drivers' ODs did not correlate with in-car glance lengths or visual short-term memory capacity. The handwriting method was further studied with 24 participants with instructions and practice on writing eyes-on-road. The practice did not affect the test results. The findings suggest that handwriting could be visually less demanding than touch screen typing but the reliability of character recognition should be improved or the driver well-experienced with the method to minimize its distraction potential.

Author Keywords
Driver distraction; visual demand; visual occlusion; occlusion distance; text entry methods; visual short-term memory; Visual Patterns Test.

CCS Concepts
• Human-centered computing~HCI design and evaluation methods • Human-centered computing~Graphical user interfaces • Human-centered computing~Interaction devices

INTRODUCTION
According to several studies, text entry with a touch screen keyboard is among the most visually distracting in-car tasks for the driver (e.g., [17,28,24]). Yet, it seems that many drivers are willing to take the risk, as it seems that short messaging with a smartphone is among the most popular in-car activities (e.g., [11]). Many in-car activities, that can support the primary task of driving, such as destination entry (way-finding) and music search (entertainment for keeping alert), may also require text entry. For these reasons, there is a need for visually less demanding in-car text entry methods than the touch screen keyboard (e.g., 29]).

AutomotiveUI '17, September 24–27, 2017, Oldenburg, Germany

ACM ISBN 978-1-4503-5150-8/17/09…$15.00
https://doi.org/10.1145/3122986.3122987

In this study, three different in-car text entry methods were compared: touch screen keyboard, handwriting and voice recognition. A voice recognition-based text entry has been shown to be significantly less distracting than a keyboard text entry in several controlled studies (e.g., [9,10]). However, as pointed out by Reimer and Mehler [22], against common belief, also the voice-guided systems typically include some visual-manual interactions, which may be distractive. Handwriting on a touch screen is a rather new method for the automotive context, and it appears that there is not yet much published research concerned with this method. For example, Kern at al. [12] studied handwritten text in the automotive context, but comparative distraction testing of handwriting as a text entry method seems to be lacking. The advantage of handwriting is that it may enable the driver to keep eyes on the road while writing especially if the system gives audio feedback to the driver, that is, repeats the written letters out loud.

According to Foley, Young, Angell and Domeyer [5, p. 62], "*visual distraction is any glance that competes with activities necessary for safe driving*". The definition of visual distraction by Foley et al. [5] is incomplete, as it does not define the "*activities necessary for safe driving*". This incompleteness places challenges for the operationalization of visual distraction. According to the study by Kircher and Ahlstrom [13] there is minimum required attention for each driving situation that can be fulfilled by different visual sampling patterns off road. This suggests that not all off-road glances are equally distractive but the timing of an off-road glance plays a critical role in visual distraction. A distracting off-road glance can be interpreted as a calibration failure between the (momentary) visual demands of driving and the individual preference for an off-road glance length, following the task-capability interface model by Fuller [6]. Here, we refer to visual distraction, in short, as a calibration failure between driving task's visual demand and the driver's off-road glance length.

Following these lines of thought, a novel distraction testing method, introduced by Kujala and Mäkelä [15] was used in our study to evaluate and compare the visual distraction potential of the three text entry methods. The testing method has been used previously to study distraction potential of audio-visual route guidance (see [14]).

The testing method is based on 97 drivers' preferred occlusion distance (OD) data mapped on the test routes [16]. The concept of occlusion distance refers to a distance that is traveled during the occluded period, that is, the distance that a driver feels comfortable to drive without visual information while concentrating on the driving task. In the testing, the median ODs of the 97-driver sample for each 1-by-1-meter test route point are utilized as a baseline for acceptable in-car glance lengths (distances, to be exact). These are labeled as *green* in-car glances. The in-car glance distances exceeding the 85th percentile of the 97-driver sample at a road point are considered as calibration failures (following Fuller, [6]) and are labeled as *red* in-car glances. A red glance suggests that the in-car task has (momentarily) caught the driver's visual attention for a longer period of time than what a great majority of drivers would not prefer to glance off road when focusing on driving at that route point.

The testing method also strives to take the drivers' individual off-road glance length preferences into account as previously has been studied that these can significantly affect the results of the distraction testing (e.g., [2]). In order to analyze the reliability and validity of the test results, we studied the test participants' individual preferences for in-car glance distances and ODs and validate the comparability of the latter with the OD distribution of the 97-driver sample. In addition, we were interested to see if the OD preference could be explained by a capability-related measure of visual short-term memory capacity (Visual Patterns Test; [4]) and if the experiences of task demands are in line with the objective distraction metrics. The specific research questions for the distraction testing were:

1. Do the studied in-car tasks pass the used verification criteria set by Kujala et al. [14]?
2. Are there significant differences in the visual distraction potential between text entry by keyboard, handwriting and voice recognition?
3. Do the drivers experience different levels of task workload between the text entry methods?

For method validation, the research questions were:

4. Are the individually preferred ODs of the test participants comparable to the ODs of the baseline sample of 97 drivers [16]?
5. Is there an individual preference threshold for the in-car glance distances across the tasks?
6. Do the driver's OD's correlate with their preferred in-car glance distances across the tasks?
7. Do the drivers' ODs correlate with their Visual Patterns Test scores [4]?

EXPERIMENT 1 - COMPARATIVE DISTRACTION TEST

The experimental design of the distraction testing was within-subjects (one IV with three levels), the independent variable being the text entry method (keyboard, handwriting, voice recognition).

Participants

The NHTSA [20] recommendations on the driver sample for testing distraction of in-vehicle electronic devices were followed as closely as possible. The participants were recruited via university's mailing lists. In total 17 participants finished the experiment, twelve males and five females. Seven female participants had to quit the experiment because of symptoms of simulator sickness. Three of them were able to complete the occlusion trial and the Visual Patterns Test, and thus, the correlation tests between these include 20 participants.

The age of the participants varied from 20 to 63 years, mean age being 34.4 years (*SD* = 12.2). Five of the participants were 18 to 24 years old, six 25 to 39 years old, four 40 to 54 years old and two were older than 55 years. All participants had a valid driver's license and all of them drove at least 5 000 kilometers per year. The total kilometers driven per year varied from 7 000 to 30 000 and with a mean of 15 352 kilometers (*SD* = 7 526) per year. The driving experience varied from two to 45 years, with a mean of 16.8 (*SD* = 12.0) years. All participants had normal vision. The experiments were instructed in Finnish and all participants were fluent in Finnish.

Apparatus

The experiments were conducted at the driving simulator laboratory at the University of Jyväskylä. The driving simulator can be described as medium-fidelity with the CKAS Mechatronics 2-DOF motion platform. The simulator consisted of longitudinally adjustable seat, Logitech G27 force-feedback steering wheel and pedals (Figure 1). During the experiments, automatic transmission was used. Three 40" LED screens (95.6 cm x 57.4 cm, resolution 1440 x 900 pixels per screen) were used to display the driving scene. A rear-view mirror, a head-up display speedometer and a RPM gauge were displayed on the middle screen. Both side screens had side mirrors. For the occlusion trial, both sides of the steering wheel were equipped with a lever that revealed the driving scene. Each press revealed the driving scene for 500 milliseconds as in the original occlusion method of Senders, Kristofferson, Levison, Dietrich and Ward [25]. Continuous pressing of the lever kept the driving scene continuously visible. The driving simulation software was provided by Eepsoft (http://www.eepsoft.fi/). Driving log data was saved at 10 Hz.

Figure 1: Experimental setup and the position of the tablet.

The predefined routes that were used during the trials simulated real Finnish suburban roads located at Martinlaakso, Vantaa. The roads were the same as used in the study of Kujala and Mäkelä [15]. The text entry methods and the in-car tasks were implemented based on Carrio application (Figure 2), an in-vehicle infotainment system (https://carrioapp.com/) running on 7" Lenovo TB3-730X tablet. In order to make a search with the keyboard, the user needed to tap the search field to activate the keyboard, type the search phrase and tap the magnifying glass key. To activate the handwriting method (developed by http://www.myscript.com/), the user had to tap the handwriting icon, enter the letters one letter at a time and finally, tap the check mark icon. The handwriting method gave audio feedback, that is, repeated the written letter out load, and thus, enabled writing without visual attention. The voice recognition search was activated by tapping the microphone icon. For all the methods, the system listed several search results to choose from by tapping the result. Ergoneers' Dikablis 50 Hz head-mounted eye-tracking system was used to record participants' eye movements and a LAN bridge was used for the synchronization of the driving simulator (x, y, speed) and the eye-tracking data.

Procedure

The demographic data was collected before the experimentation via email. The participants signed an informed consent form before participating. Before the actual experiment, participants practiced driving in an artificial city environment with other road users. They were instructed to drive as long as they wanted, with an average practice time of 3.0 minutes. After they felt comfortable with driving, they started practicing for the occlusion drive: how to drive vision occasionally occluded and how to use the levers that removed the occlusion and revealed the driving scene. The practice took place in a same city environment with other road users as the previous practice. Mean practice time was 3.65 minutes.

Figure 2: Three different text entry user interfaces: keyboard (on top), handwriting and voice recognition.

After the practices, the experiment started with the occlusion trial for test sample validation. In the trial, the screens were blank by default and the participants were able to see the driving scene for 500 milliseconds by pressing the levers on the steering wheel. In the trial, the participants were instructed to follow the traffic rules, to drive safely and at the same time to drive without visual information (i.e. vision occluded) as long as possible. An extra movie ticket was promised to those six drivers who could drive the longest periods without visual information but still accurately. This was done in order to make the participants to focus on the driving task but still trying to maximize the occlusion distance to their preference. A highway route without other road users was used in the occlusion trial. The route was the same as for the baseline sample of $N = 97$ in Kujala et al. [16]. The speed limits varied from 60 to 80 to 120 kilometers per hour during the trial and each change in a limit was given at the same point of the route by the experimenter. After the trial, each participant filled out the NASA-TLX questionnaire [7].

After the occlusion trial, the Visual Patterns Test [4] was completed. Once the test was done, the eye-tracking headset was put on, adjusted and calibrated and then the distraction test part started. In the distraction testing, the participants were instructed to prioritize the driving task, to obey the traffic regulations and to drive safely. The speed limit was set to 50 kilometers per hour, but the participants were able to adjust the speed freely if needed.

Each participant completed three different tasks with three different text entry methods: keyboard, handwriting and voice recognition. The tasks were:

1. To write and find three different addresses
2. To write and start to play three different songs
3. To write and find three different contact information.

All tasks were completed with each text entry method. The order of the tasks and the driven routes were counterbalanced in order to avoid learning effect. The visual demands of the used routes were as similar as possible (as measured by OD in [15]) and there were no other road users on the routes. The traffic lights were always green when participant approached to junctions. After each text entry method tasks, the NASA-TLX questionnaire was filled out. Every participant was rewarded with a movie ticket after the experiment.

Analyses

The main dependent variables in the distraction testing were the percentage of red in-car glances (in-car glance distances above 85th percentile ODs for the 1x1-meter route points), the percentage of green in-car glances (in-car glance distances below or at median ODs for the 1x1-meter route points) and reduced NASA-TLX (no weighting) for each text entry method. In addition, we compared the number of the in-car glances and the number of errors (i.e. incorrectly recognized input and typing errors for keyboard) per text entry method. Drivers' occlusion distances (m) and in-car glance distances (m, distance traveled during an in-car glance) were measured for sample validation.

The in-car glance lengths were scored in real-time by a script that read the pupil's x and y coordinates from the eye-tracker. The coordinates were synchronized with the location data that the driving simulator provided. The glance lengths were scored following the SAE-J2396 [26] definition, with the exception that the gaze transition time back to the driving scene was added to a glance, in order to enable more direct comparability with the occlusion distance (no focal visual information available from the road during an in-car glance). All glances were manually searched from a synchronized video (25 fps) for validity using Noldus Observer XT software. All inaccuracies were manually corrected frame-by-frame.

The verification threshold for the red glances was set to 6 % and to 68 % for the green glances, based on Kujala et al. [14]. The percentage thresholds are based on the median percentages of the occlusion distances below or at the median OD ('green occlusions') and the occlusion distances exceeding the 85th percentile OD ('red occlusions') of the 97 drivers in Kujala et al. [16]. To test the equality of the median red and green in-car glance percentages of the three text entry methods to the verification thresholds (6 % and 68 %), one-sample sign test was used due to the highly non-Gaussian distributions. The differences between the text entry methods were analyzed with Wilcoxon signed-rank test. In order to test the differences in the experienced task workload between the three text entry method trials and the occlusion trial, one-way repeated measures ANOVA was used. When the sphericity assumption was violated, the Greenhouse-Geisser correction was applied. Bonferroni corrections were applied for pairwise comparisons. Cronbach's Alpha was used to test the correlation and covariance between in-car glance distances across the different text entry tasks.

In the occlusion trial, the dependent variable was occlusion distance (OD). Because the occlusion metrics were non-Gaussian, median was used instead of mean. In order to control the effects of accelerations and decelerations in the beginning, in the intersections and in the end of the trial, only occlusion distances that were driven over 20 m/s (72 km/h) were included in the data. The Pearson product-moment correlation was used to test the correlation between median occlusion distance and median in-car glance distance as well as the correlation between median occlusion distance and the Visual Patterns Test scores [4]. The equality of the test drivers' OD distribution to the OD distribution of $N = 97$ in Kujala et al. [16] was assessed by Levene's test of equality of variances.

EXPERIMENT 1 - RESULTS

Number of glances and errors

Mean number of in-car glances during the nine tasks per method was 120 ($SD = 31$) for the keyboard, 201 ($SD = 61$) for the handwriting, and 84 ($SD = 26$) for the voice recognition. All the differences were significant ($p < .001$). There was a significant difference between the number of errors in the keyboard ($M = .9$, $SD = 1.3$) and the handwriting ($M = 4.2$, $SD = 2.5$) tasks ($Z = 3.304$, $p = .001$) as well as between the handwriting and the voice recognition ($M = 1.2$, $SD = 1.1$) tasks ($Z = 3.225$, $p = .001$). No difference was found between the keyboard and the voice recognition tasks ($p = .298$).

Red in-car glances

The verification threshold of the red glances was set to 6 % (at or below). The keyboard tasks did not pass the verification criteria, one-sample sign test indicating that the percentage of red glances was significantly higher than 6 % ($p = .003$, median = 13.22 % (Figure 3). Either did the handwriting tasks, the percentage of red glances being also significantly higher than 6 % ($p = .001$, median = 9.49 %). The voice recognition tasks passed the verification criteria, the median percentage being 3.51 % ($p = .722$). Wilcoxon signed-rank test indicated that there was no difference in the percentages of red glances between the keyboard and the handwriting tasks ($Z = 1.349$, $p = .177$). However, there was a significant difference between the keyboard and the voice recognition tasks ($Z = 3.337$, $p = .001$) as well as the handwriting and the voice recognition tasks ($Z = 2.864$, $p = .004$).

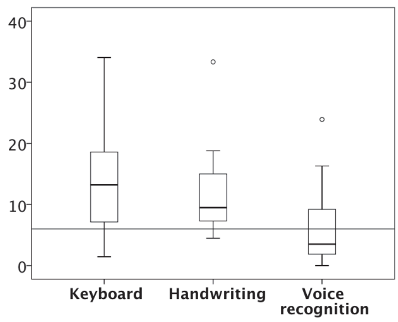

Figure 3: Percentage of red in-car glances per text entry method (the verification threshold at 6 %).

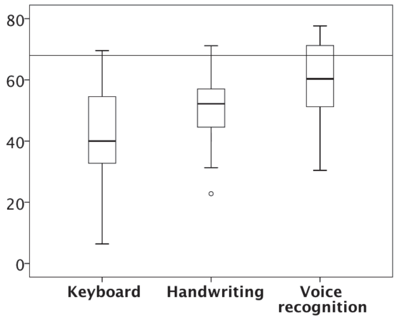

Figure 4: Percentage of green in-car glances per text entry method (the verification threshold at 68 %).

Green in-car glances

The verification threshold of green glances was set to 68 % (at or above). Only the voice recognition tasks passed the verification criteria, by not differing significantly from 68 % (Figure 4). The median of green in-car glances in the voice recognition tasks was 60.35 % (p = .055) whereas in the keyboard tasks median was 40.00 % (p < .001) and in the handwriting tasks 52.20 % (p < .001). After Bonferroni correction, there was no significant difference in the percentages of green glances between the keyboard and the handwriting tasks (Z = 1.965, p = .049, α = .017). However, there was a significant difference between the keyboard and the voice recognition tasks (Z = 3.432, p = .001) as well as the handwriting and the voice recognition tasks (Z = 2.580, p = .010).

Experienced task workload - NASA-TLX

A significant main effect of trial was found, F(2.174, 38.783) = 12.819, p < .001, partial η^2 = .445. The handwriting tasks were experienced more demanding than the keyboard (mean difference 15.44, p < .001) and voice recognition tasks (mean difference 22.55, p < .001, Figure 5). After Bonferroni correction, the difference between the handwriting tasks and the occlusion trial was not significant (p = .031, α = .0083). The difference between the keyboard tasks and the occlusion trial was also not significant (p = .104). The voice recognition tasks were experienced significantly less demanding than the occlusion trial (p = .006).

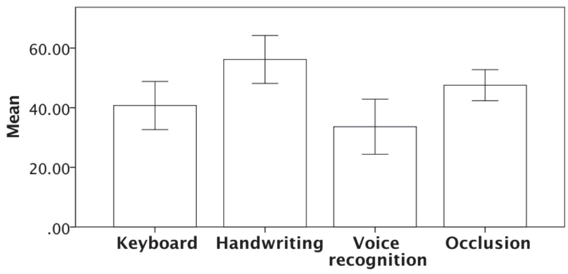

Figure 5: Experienced workload (NASA-TLX, max 100). Error bars: 95% CI.

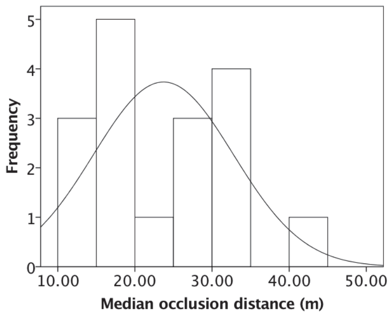

Figure 6: Individual median occlusion distances (m) on highway (speed > 72 km/h)

Occlusion distances

For sample validation, the distributions of the drivers' median ODs (Figure 6) were compared to the median ODs of the baseline data (N = 97; [16]). Drivers' median OD varied from 11.3 to 43.0 meters (median of 21.5 m). Levene's test indicated that the variance of the OD distribution does not differ significantly from the baseline OD distribution of N = 97 in Kujala et al. [16] (F = 1.07, p = .303) with a range between 3.2 to 41.9 meters.

In-car glance distances across the tasks, occlusion distance and visual short-term working memory

With each text entry method, three different types of tasks (3x3 tasks) were conducted: entering an address, entering a song, and entering a contact entry. High correlations between the 9 tasks were found and the Cronbach's alpha was excellent (α = .901). However, there was no correlation between the occlusion distances and the in-car glance distances (r = -.002, p = .994). No correlation was found either between median occlusion distances and Visual Patterns Test scores (N = 20, r = .232, p = .324, [4]).

EXPERIMENT 1 - DISCUSSION

The visual distraction potential of three different text entry methods was studied following the testing and verification criteria of Kujala and Mäkelä [15]. Only the voice recognition-based text entry tasks passed the set verification criteria. The percentage of red in-car glances during the voice recognition tasks (3.51 %) was significantly lower than the verification threshold of 6 % as well as that of the keyboard (13.22 %) or handwriting (9.49 %) tasks.

Previous studies have shown similar results concerning the differences between voice recognition and touch screen keyboards (e.g., [3,9,23,27,28]).

The experienced task workload was the highest for the handwriting tasks, even higher than for the occlusion trial. The novelty of the handwriting method as well as the number of recognition errors by the system could explain some of the experienced workload. However, the method shows some promise as the percentage of red glances stayed at the same or even lower level than that for the keyboard tasks, even if there were significantly more errors (and thus, more glances) for the former. With higher recognition accuracy and more experienced users the method could be visually significantly less demanding than keyboard text entry. The voice recognition tasks were experienced as least demanding of all the tasks. During these tasks, manual input was considerably less needed than during the keyboard or the handwriting tasks.

The distribution of the occlusion distances (OD) was fairly similar to the baseline data [16] and the differences in the individual OD preferences can be assumed to not have affected the results of the distraction testing. Drivers with low ODs are not over-presented in the sample compared to the baseline data. The results of the distraction testing are well in line with earlier test results on significantly less demanding audio-visual route guidance (0.0-2.5 % red glances, Kujala et al., [14]).

The individual in-car glance length preference was found across the different in-car tasks as in previous studies [1,2,14,18,21]. However, correlation between the occlusion distances and the in-car glance distances was not found. In Kujala et al. [14] a correlation was found, but the locations of the in-car glances were more controlled (to follow route guidance) and the sample size was N = 24, here N = 17. In future studies, a more accurate metric to study this association would be, for instance, the ratio between in-car glance distance and the median OD of the baseline data (N = 97) on the route point where the glance is started. This would control the variability of the visual demands on the route points where the in-car glances are initiated. No correlation between OD and short-term visual memory capacity was found. Again, the sample size (N = 20) was small but it is unlikely there is more than a weak association between these two measures. More research is needed in order to explain the individual OD and in-car glance distance preferences.

In order to test our hypothesis on the effects of experience on eyes-on-road text entry with the handwriting method, we conducted another experiment focusing on this method. In addition, we wanted to study further the relationship between ODs and in-car glance lengths in text entry tasks.

EXPERIMENT 2 - INSTRUCTED HANDWRITING

In Experiment 2 we hypothesized that the handwriting method would have significantly lower visual distraction potential if the drivers would have practiced the use of the method without vision. Again, the handwriting method gave audio feedback, that is, repeated the written letter out load after each entry. We studied the question with 24 new participants, and compared the test results to those of Experiment 1.

Participants

The NHTSA [20] recommendations on the driver sample were followed as accurately as possible. The recruitment of the participants was done via university's mailing lists. In total 24 participants took part in the experiment: 17 males and 7 females. Five women indicated symptoms of simulator sickness and were replaced with male participants.

The age of the participants ranged from 20 to 79 years, mean age being 34.8 years (SD = 16.0). Eight of the participants were 18 to 24 years old, nine 25 to 39 years old, four 40 to 54 years old and three were older than 55 years. All participants had a valid driver's license and drove at least 5 000 kilometers per year. The total kilometers driven per year varied from 5 000 to 30 000, with a mean of 12 938 kilometers (SD = 7 046) per year. Their driving experience varied from 2 to 55 years, with a mean of 16.0 (SD = 15.0) years. Normal or corrected-to-normal vision was a prerequisite for participating. The experiments were instructed in Finnish and all participants understood and spoke Finnish. The participants were rewarded with a gift certificate (15 EUR) for participating the study.

Apparatus

The experiments were conducted at the driving simulator laboratory at the University of Jyväskylä and the used apparatus was the same as in Experiment 1. The used routes during the trials simulated real Finnish highways located at Martinlaakso, Vantaa and were the same as the ones used in the study of Kujala et al. [16]. This time, highway routes were used in order to keep the environmental visual demands of the driving as static and similar as possible for the analysis of the association between ODs and in-car glance distances. During the trials, no other road users were on the routes.

Procedure

The demographic data was collected in advance via email. An informed consent form was signed before the experiment. The practices were conducted similarly as in Experiment 1. The mean driving practice time was 5.79 minutes and the mean occlusion trial practice time was 4.33 minutes. The experiment started with the occlusion trial for test sample validation. The instructions were exactly the same as in Experiment 1. After the occlusion trial, NASA-TLX questionnaire [7] was filled out, the eye-tracking headset was put on, adjusted and calibrated and the distraction testing for the handwriting task started.

The participants were shown how the handwriting method is applied and how to write without glancing at the tablet's screen (see Figures 1 and 2 [middle]). After the demonstration was their turn to repeat the exercise and rehearse to write without glancing at the screen (simulator stationary). The experimental task was to write an address using the handwriting method. The participants received an additional instruction to try to avoid glancing at the screen while writing and driving. The nominal speed limit changed from 120 to 80 kilometers per hour in the middle of the route after changing the road via junction, and this change was told to each participant at the same point of the route. They were also advised that they can freely adjust the speed if necessary. The route was a highway route with no other road users and every participant drove the same distance from the starting point to the ending point. During the drive, they wrote as many address entries as they could but in practice, two turned out to be the maximum number of addresses that a participant was able to finish. After the trial, the NASA-TLX [7] questionnaire was filled out and they were rewarded with a gift certificate.

Analysis

The main dependent variables in the distraction testing were the same as in Experiment 1. In addition, visual demand ratio, the ratio between in-car glance distance and the median OD of the baseline data (N = 97, Kujala et al, [16]) on the route point where the in-car glance is started, was measured. All the statistical analyses were conducted in the same manner as in Experiment 1 (for a single condition).

EXPERIMENT 2 - RESULTS

Number of glances and errors

The mean number of glances during the handwriting task per participant was 44 (SD = 23) and the mean number of errors per participant was 3.5 (SD = 1.9).

Red in-car glances

Again, the handwriting task did not pass the set verification criteria for the red in-car glances (Figure 7). One-sample sign test indicated that the percentage of red glances was significantly higher than 6 % (p = .036, median = 9.00 %).

Green in-car glances

The handwriting task did not pass the verification criteria for the green in-car glances (Figure 8). The percentage of the green glances was significantly lower than 68 % (p < .001, median = 52.50 %).

Experienced task workload - NASA-TLX

Wilcoxon signed rank test indicated that the handwriting task (M = 57.29, SD = 12.17) was experienced significantly more demanding than the occlusion trial (M = 49.93, SD = 12.92, Z = -2.173, p = .030, d = .587).

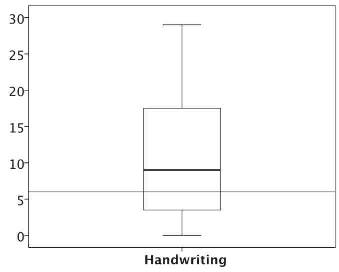

Figure 7: Percentage of red in-car glances (the verification threshold at 6 %).

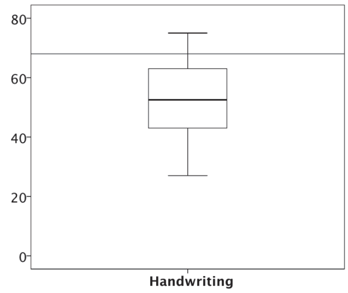

Figure 8: Percentage of green in-car glances (the verification threshold at 68 %).

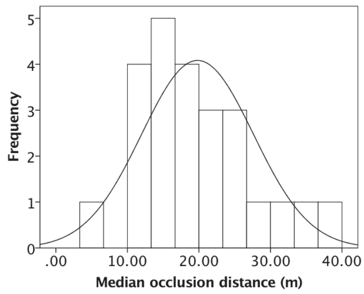

Figure 9: Individual median occlusion distances (m) on highway (speed > 72 km/h)

Occlusion distances

The median OD of the drivers varied from 6.35 to 37.18 meters, median being 17.98 meters (Figure 9). Levene's test indicated that the variance of the OD distribution does not differ significantly from the baseline OD distribution of N = 97 in Kujala et al. [16] (F = .08, p = .778). No correlation was found neither between OD and in-car glance distance (r = -.193, p = .366) nor between OD and visual demand ratio (r = -.284, p = .179). A strong correlation between in-car glance distance and visual demand ratio was found, r = .824, p < .001.

EXPERIMENT 2 - DISCUSSION

The visual distraction potential of the handwriting text entry method was re-evaluated with 24 drivers getting practice and instructions on eyes-on-road writing. Surprisingly, the results were highly similar to the findings in the first experiment. In the Experiment 1, the percentage of the red glances was 9.49 % and in Experiment 2, the percentage was 9.00 %. The percentage of the green glances in the first experiment was 52.20 % and in the second experiment 52.50 %. The percentages indicate that the handwriting as a text entry method did not pass the set verification criteria in either experiment. It seems that the rather short practice of writing without glancing at the screen did not work that effectively for minimizing the tasks' visual distraction potential. Perhaps longer experience in using the handwriting method could improve the skill to write without watching the screen while driving. The relatively high number of character recognition errors could also have affected the percentage of the red in-car glances (Experiment 1: mean 4.2, Experiment 2: mean 3.5). The number of errors during the handwriting task was still quite high in Experiment 2 despite of the practice to write eyes-on-road.

As previously, the experienced task workload was higher during the handwriting task than during the occlusion trial. Again, we assume that the high number of recognition errors led to high levels of experienced task workload due to higher visual demand of the task in the form of additional glances for making corrections. Predictive text input [19,29], allowing for more inaccurate input for individual characters, could significantly decrease the visual distraction potential of the method but this should be further studied. A limitation of the used testing environment is the degrees of freedom in the movements of the motion platform (2 DOF). Road surface roughness and other vertical movements of the vehicle, which could further affect the usefulness of the handwriting method (as well as touch screen keyboard) on real roads, were absent in the driving simulation.

The distribution of the occlusion distances (OD) was again fairly similar in Experiment 2 as in the baseline data of Kujala et al. [16], probably due to the inclusion of different age groups (older drivers preferring shorter ODs, $r = -.437$, $p = .037$, $N = 23$). It can yet again be assumed that the individual OD preferences did not affect the results of the distraction testing. Most importantly, low OD drivers were not over-represented in the sample compared to the baseline data or the sample in Experiment 1.

The effects of the varying visual demands were better controlled in Experiment 2 than in Experiment 1 due to the highway routes. This is evident from the strong correlation between the in-car glance distances and the visual demand ratio, indicating that the visual demands did not vary significantly between the in-car glances. Yet again, OD did not correlate with in-car glance lengths. We assume that the missing correlation is due to variability in the participants' capabilities in the writing task and the nominal speed limits. The low-OD drivers (aged, in particular) may have required longer in-car glances than the more skilled writers but they were not able to compensate sufficiently for this by decreasing driving speed due to the instructed speed limits [6]. Secondary task related skills and also the structural constraints (e.g., natural break points, [8]) of an in-car task might affect the in-car glance lengths more than the individual uncertainty that may rarely rise to the same level with these kinds of in-car tasks than in an occlusion experiment (cf. the percentages of green in-car glances). An open question remains, if the sample should be validated based on the OD or the in-car glance length distributions. As the metric of red in-car glances is based on ODs and it seems to provide reliable results, we suggest the former is more important for this type of distraction testing.

The highly similar distraction test results for the handwriting tasks between Experiment 1 and Experiment 2 provide reliability for the suggested testing method and verification criteria. The test results were highly similar even if the participant sample, the road environment (suburban vs. highway), and driving speeds were different. This suggests that comparable test data could be gathered with the testing method even if it is applied to different driving simulator and driving scenario implementations, in which the baseline occlusion data can be collected.

CONCLUSIONS

The visual distraction potential of three different text entry methods was studied following the testing and verification criteria of Kujala and Mäkelä [15]. Only the voice recognition-based text entry tasks passed the set verification criteria based on the percentages of red and green in-car glances (in-car glance lengths above 85th percentile or below median of the baseline ODs, correspondingly). The percentage of red in-car glances during the voice recognition tasks (3.51 %) was significantly lower than the verification threshold of 6 % as well as that of the keyboard (13.22 %) or handwriting (9.49 %) tasks. The voice recognition tasks were also experienced as least demanding of all the tasks.

The handwriting method was further studied with 24 participants with instructions and practice on writing with eyes on road. The practice on the method did not affect the test results significantly. The findings suggest that handwriting could be visually less demanding than touch screen typing but the reliability of the text input recognition should be significantly improved or the driver well-experienced with the method in order to minimize its visual distraction potential. The handwriting method could be further researched with participants who are already familiar with using the method.

The highly similar distraction test results for the handwriting tasks between Experiment 1 and Experiment 2 provide reliability for the suggested testing method and verification criteria.

ACKNOWLEDGEMENTS

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

REFERENCES

1. Robert Broström, Peter Bengtsson and Mikael Ljung Aust. 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. http://dx.doi.org/10.1016/j.trf.2015.10.017
2. Robert Broström, Mikael Ljung Aust, Linnea Wahlberg and Laban Källgren. 2013. What drives off-road glance durations during multitasking–capacity, practice or strategy. In *Proceedings of the 3rd International Conference on Driver Distraction and Inattention*.
3. Joseph M. Crandall and Alex Chaparro. 2012. Driver distraction: Effects of text entry methods on driving performance. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 56,1: 1693-1697. https://doi.org/10.1177/1071181312561339
4. Sergio Della Sala, Colin Gray, Alan Baddeley, Nadia Allamano and Lindsey Wilson. 1999. Pattern span: a tool for unwelding visuo–spatial memory. *Neuropsychologia*, 37, 10: 1189-1199. http://dx.doi.org/10.1016/S0028-3932(98)00159-6
5. James P. Foley, Richard Young, Linda Angell and Joshua E. Domeyer. 2013. Towards operationalizing driver distraction. In *Proceedings of the 7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, 57-63.
6. Ray Fuller. 2005. Towards a general theory of driver behaviour. *Accident Analysis & Prevention* 37, 3: 461-472. http://dx.doi.org/10.1016/j.aap.2004.11.003
7. Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, 52, 139-183. https://doi.org/10.1016/S0166-4115(08)62386-9
8. Christian P. Janssen, Duncan P. Brumby and Rae Garnett. 2012. Natural break points: The influence of priorities and cognitive and motor cues on dual-task interleaving. *Journal of Cognitive Engineering & Decision Making*, *6*, 5-29. https://doi.org/10.1177/1555343411432339
9. Jibo He, Alex Chaparro, Bobby Nguyen, Rondell J. Burge, Joseph Crandall, Barbara Chaparro, Rui Ni and Shi Cao. 2014. Texting while driving: Is speech-based text entry less risky than handheld text entry? *Accident Analysis & Prevention*, 72, 287-295. http://dx.doi.org/10.1016/j.aap.2014.07.014
10. Jibo He, William Choi, Jason S. McCarley, Barbara S. Chaparro and Chun Wang. 2015. Texting while driving using Google Glass™: Promising but not distraction-free. *Accident Analysis & Prevention*, 81, 218-229. http://doi.org/10.1016/j.aap.2015.03.033
11. Petri Jääskeläinen and Leena Pöysti. 2014. *Tarkkaamattomuus tieliikenteen turvallisuusongelmana – suomalaisten käsityksiä* (Driver distraction and inattention as a safety issue in road traffic – Finnish views). Helsinki: Finnish Road Safety Council.
12. Dagmar Kern, Albrecht Schmidt, Jonas Arnsmann, Thorsten Appelmann, Nillakshi Pararasasegaran and Benjamin Piepiera. 2009. Writing to your car: handwritten text input while driving. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*, 4705-4710. ACM. 10.1145/1520340.1520724
13. Katja Kircher and Christer Ahlstrom. (in press). Minimum required attention: a human-centered approach to driver inattention. *Human factors*. https://doi.org/10.1177/0018720816672756
14. Tuomo Kujala, Hilkka Grahn, Jakke Mäkelä and Annegret Lasch. 2016a. On the Visual Distraction Effects of Audio-Visual Route Guidance. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 169-176. ACM. 10.1145/3003715.3005421
15. Tuomo Kujala and Jakke Mäkelä. 2015. Development of a testing environment and a verification procedure for in-car tasks with dynamic driving scenarios. In *Proceedings of the 4th International Conference on Driver Distraction and Inattention*.
16. Tuomo Kujala, Jakke Mäkelä, Ilkka Kotilainen and Timo Tokkonen, 2016b. The Attentional Demand of Automobile Driving Revisited: Occlusion Distance as a Function of Task-Relevant Event Density in Realistic Driving Scenarios. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 58, 1: 163-180. https://doi.org/10.1177/0018720815595901
17. Tuomo Kujala, Johanna Silvennoinen and Annegret Lasch. 2013. Visual-manual in-car tasks decomposed: text entry and kinetic scrolling as the main sources of visual distraction. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications,* 82-89. ACM. 10.1145/2516540.2516562
18. Mikael Ljung Aust, Sergejs Dombrovskis, Jordanka Kovaceva and Bo Svanberg. 2013. An empirically based suggestion for reformulating the glance duration criteria in NHTSA's visual-manual interaction guidelines. *SAE International Journal of Passenger Cars-Electronic and Electrical Systems*, 6(2013-01-0444), 444-453. 10.4271/2013-01-0444
19. I. Scott MacKenzie and R. William Soukoreff. 2002. Text entry for mobile computing: Models and methods, theory and practice. *Human–Computer*

Interaction, 17, 2-3: 147-198. 10.1080/07370024.2002.9667313

20. National Highway Traffic Safety Administration. (2013). *Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices*. (NHTSA-2010-0053.) Washington DC, NHTSA.
21. Jami Pekkanen, Otto Lappi, Teemu H. Itkonen and Heikki Summala. 2017. Task-difficulty homeostasis in car following models: experimental validation using self-paced visual occlusion. *PLoS one*, 12, 1: e0169704. http://dx.doi.org/10.1371/journal.pone.0169704
22. Bryan Reimer and Bruce Mehler. 2013. The effects of a production level "voice-command" interface on driver behavior: summary findings on reported workload, physiology, visual attention, and driving performance. MIT AgeLab Technical Report No. 2013-17A. Massachusetts Institute of Technology, Cambridge, MA.
23. Bryan Reimer, Bruce Mehler, Ian Reagan, David Kidd and Jonathan Dobres. 2016. Multi-modal demands of a smartphone used to place calls and enter addresses during highway driving relative to two embedded systems. *Ergonomics*, 59, 12: 1565-1585. http://dx.doi.org/10.1080/00140139.2016.1154189
24. Aubrey Samost, David Perlman, August G. Domel, Bryan Reimer, Bruce Mehler, Alea Mehler, Jonathan Doblers and Thomas McWilliams. 2015. Comparing the Relative Impact of Smartwatch and Smartphone Use While Driving on Workload, Attention, and Driving Performance. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 59, 1: 1602-1606. Sage CA: Los Angeles, CA: SAGE Publications. https://doi.org/10.1177/1541931215591347
25. John W. Senders, A. B. Kristofferson, W. H. Levison, C. W. Dietrich and J. L. Ward. 1967. The attentional demand of automobile driving. *Highway research record*, 195.
26. 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.
27. Kathryn G. Tippey, Elayaraj Sivaraj, Wil-Johneen Ardoin, Trey Roady and Thomas K. Ferris. 2014. Texting while driving using Google Glass: Investigating the combined effect of heads-up display and hands-free input on driving safety and performance. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58, 1: 2023-2027. Sage CA: Los Angeles, CA: SAGE Publications. https://doi.org/10.1177/1541931214581422
28. Omer Tsimhoni, Daniel Smith and Paul Green. 2004. Address entry while driving: Speech recognition versus a touch screen keyboard. *Human Factors*, 46, 4: 600-610. https://doi.org/10.1518/hfes.46.4.600.56813
29. Gabriela Villalobos-Zúñiga, Tuomo Kujala and Antti Oulasvirta. 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,* 177-184. ACM. 10.1145/3003715.3005453

V

VISUAL DISTRACTION EFFECTS BETWEEN IN-VEHICLE TASKS WITH A SMARTPHONE AND A MOTORCYCLE HELMET-MOUNTED HEAD-UP DISPLAY

by

Hilkka Grahn & Tuomo Kujala, 2018

In Proceedings of the 22nd International Academic Mindtrek Conference, 153–162

https://doi.org/10.1145/3275116.3275134

Visual Distraction Effects between In-Vehicle Tasks with a Smartphone and a Motorcycle Helmet-Mounted Head-Up Display

Hilkka Grahn†
Faculty of Information Technology
University of Jyväskylä
Jyväskylä Finland
hilkka.grahn@jyu.fi

Tuomo Kujala
Faculty of Information Technology
University of Jyväskylä
Jyväskylä Finland
tuomo.kujala@jyu.fi

ABSTRACT

Besides motorists, also motorcyclists need safer user interfaces to interact with useful applications on the road. In this paper, distraction effects of in-vehicle tasks conducted with a head-up display (HUD) for motorcyclists were compared to smartphone tasks with 24 participants in a driving simulator.

Compared to the smartphone tasks, the head-up display tasks decreased the percentage of inappropriately long glances by 45 percent. The head-up display tasks were also experienced as less demanding than the smartphone tasks. Additionally, the use of head-up display for navigation did not lead to gaze concentration effects compared to baseline driving.

The head-up display is concluded to be a safer option for the tested tasks for motorcyclists than a smartphone. Based on earlier research, we assume that the use of peripheral vision allowed drivers to better maintain situational awareness during the head-up display tasks compared to the head-down smartphone tasks. In addition, the easy-to-learn haptic design of the head-up display handlebar controller could be used without vision.

CCS CONCEPTS

• Human-centered computing~User studies • Human-centered computing~Laboratory experiments • Human-centered computing~Touch screens • Human-centered computing~Haptic devices • Human-centered computing~Empirical studies in HCI

KEYWORDS

Driver distraction, visual demand, visual occlusion, occlusion distance, head-up display, head-down display, head-mounted display.

Mindtrek 2018, October 10–11, 2018, Tampere, Finland

ACM ISBN 978-1-4503-6589-5/18/10…$15.00
https://doi.org/10.1145/3275116.3275134

1 Introduction

Driver distraction and especially visual distraction has been extensively studied in recent years. Several studies have shown that visually demanding in-vehicle tasks cause visual distraction and therefore are associated with high risk of safety-critical incidents, such as near crashes and crashes (e.g., [6, 12, 26, 33]).

Existing research in this field has been limited to study visual distraction effects of in-car tasks and user interfaces (UIs) for car drivers (e.g., [1, 13, 22]). However, not only car drivers but also motorcycle drivers need, for instance, navigating aids while driving. Therefore, this study focuses on motorcycle drivers and the visual distraction associated with in-vehicle devices that are used by motorcyclists.

Compared to driving a car, driving a motorcycle is even more complex task that requires great motor skills and coordination [24]. Also, motorcyclists are one of the most vulnerable road user group. For example, in 2000, they made up less than one percent of the road traffic in the UK but suffered 14 percent of deaths and serious injuries [5].

Truong, De Gruyter and Nguyen [32] found out that motorcyclists use smartphones while driving to call, text and find information. Also, Phommachanh, Ichikawa, Nakahara, Mayxay and Kimura [27] reported that motorcyclists dial, receive calls and send text messages while operating a motorcycle.

When we add the reported phone usage to the complexity of motorcycle driving, while knowing that motorcyclists are in great risk in general in traffic, the consequences of distraction can be serious. Because of this, it is important to study the visual distraction potential of in-vehicle devices that are designed for motorcyclists and to design better user interfaces also for them to access safer the services they need on the road. It is also important that motorcyclists experience the user interfaces – designed for them – easy to use, in order to make them prefer the safer UIs over smartphones.

In this paper, we study tasks conducted with a novel motorcycle helmet-mounted head-up display (HUD) manufactured by Nuviz Inc. (https://www.ridenuviz.com/) and compare those to similar tasks with identical goals conducted with a Samsung Galaxy A3 smartphone. A novel distraction testing method introduced by Kujala and Mäkelä [17] – that categorizes in-vehicle glances to appropriate or inappropriate glances dependent on the visual demand of the road point – was used to assess the visual distraction potential of the tasks. According to several studies (see Related work), HUDs can be less distracting for drivers than head-down displays (HDD). However, HUD could cause attention capture and gaze concentration effects [34]. With this in mind, we measure also the horizontal gaze activity of the drivers during the HUD navigation tasks.

Accordingly, the three research questions for the study were:

RQ 1: Are there significant differences in the visual distraction potential between in-vehicle tasks conducted with the Nuviz head-up display and with the Android smartphone?

RQ 2: Are there significant differences in experienced workload between the Nuviz head-up display and the Android smartphone tasks?

RQ 3: Are there gaze concentration effects in the Nuviz head-up display tasks compared to baseline driving?

2 Related work

2.1 Head-up displays, head-down displays and head-mounted displays

Head-up displays for in-vehicle use have been studied previously as well but this is the first study that compares distraction effects between a HUD designed for motorcyclists and a smartphone. Head-up displays may have significant potential for reducing visual distraction by in-vehicle tasks compared to head-down displays (HDD). For example, Weinberg, Harsham and Medenica [36] compared three systems – head-down display, head-up display and auditory display – for presenting textual lists. They found out that the number of in-vehicle glances was doubled when using HDD compared to HUD. In their study, the HUD was operated with a steering wheel-mounted controller.

Villalobos-Zúñiga, Kujala and Oulasvirta [35] studied a text entry method that consisted of a physical 3x4 keypad in the steering wheel and a HUD and compared it to a touchscreen QWERTY keyboard in the center console. The results showed that the physical keypad and HUD combination allowed drivers to maintain more visual attention on the road (up to 64 %). There were also less lane deviations when using the HUD combination compared to a touchscreen keyboard.

Smith, Gabbard and Conley [29] also compared HDD and HUD. During the experiments, participants were required to conduct visual search tasks while driving. They noticed that performing tasks with HUD caused less severe decrements in driving performance than with HDD. Interestingly, they also found out that HUD affected more negatively the glance patterns on the NHTSA metrics [25] than HDD. Since there were no significant decrements on driving performance while using HUD, they concluded that people could use different visual search methods with HUDs than with HDDs. Therefore, the NHTSA guidelines [25] may not be the best practice to assess the visual distraction of in-vehicle HUDs.

Lauber, Böttcher and Butz [21] compared HUDs and head-mounted displays (HMD). HMDs have the same features as HUDs but because they are head-mounted, the information showed is always available regardless of head position. Lauber et al. [21] concluded that their study could not show any significant differences in driving performance between the tested interaction techniques. In this study, the Nuviz HUD is helmet-mounted, and thus the tested HUD is also an HMD.

Smith, Streeter, Burnett and Gabbard [30] point out that HUD interfaces should be carefully designed. HUD tasks that do not support resumability may cause even greater problems than traditional head-down displays. In general, it is wise to remember that HUDs and HMDs can also be distracting and impair driving performance [9, 10]. Even if HUDs and HMDs may allow the driver to keep peripheral vision available to serve the goals of the driving task, these may cause negative gaze concentration effects [34] compared to driving without any secondary tasks.

2.2 Visual distraction - operationalization and measurement

For instance, Foley, Young, Angell and Domeyer [7] (p. 63) have defined visual distraction as follows: "*Visual distraction is any glance that competes with activities necessary for safe driving*". Among others, Klauer, Dingus, Neale, Sudweeks and Ramsey [12] have reported in their naturalistic driving study that over two seconds eyes-off-road durations – that is, visual distraction – were associated with near-crash and crash risk.

In this study, we used a novel distraction testing method that was introduced by Kujala and Mäkelä [17], to evaluate the visual distraction potential of the Nuviz HUD tasks and the smartphone tasks. One benefit of the method is that the in-vehicle glances made during the testing can be categorized to appropriate (green) or inappropriate (red) glances dependent on the visual demands of the road point where the glance begins.

The method follows the idea of Victor et al. [33] regarding the high change rate of the driving situation and Kircher and Ahlstrom's [11] idea about the timing of the off-road glance. Victor et al. [33] have noticed that many crashes occur because of a combination of a relatively short glance and high rate at which the dynamics of the driving situation changed during that glance – not because the off-road glance was too long as such. Also, Kircher and Ahlstrom [11] have suggested that all off-road glances are not equally distracting but timing of the off-road glance is critical.

In this study, glances towards the HUD view are interpreted as off-road (i.e., in-vehicle) glances since at least the driver's focal visual attention is then on the HUD. The appropriateness of an in-vehicle glance is determined here based on the visual demands of the route point and not solely based on the glance duration. Thus, the method may be more suitable to assess if the HUD affects drivers' situational awareness and glance timing than the NHTSA [25] recommended practice with static driving scenario and static glance acceptance criteria (see [29]).

HUD technology could cause a phenomenon called tunnel vision where the gaze is concentrated on too narrowly to the HUD and/or the road center [34], sacrificing observations for unexpected events in the road environment. Because of this, it is important to study also gaze concentration effects as a form of visual distraction particularly significant for HUDs. Victor, Harbluk and Engström [34] used a metric called percent road centre to measure these types of effects with in-vehicle tasks. This metric can tell if driver's field of view decreases compared to baseline driving. In this study, we analyze the gaze concentration effects by standard deviation of gaze in x-coordinates, which is another traditional measure used for analyzing differences in visual search patterns between novice and experienced drivers (e.g., [4]). The driving scenarios we utilized include turns in junctions, which stress the importance of horizontal observations for crossing traffic.

3 Method

For measuring the visual distraction potential of the in-vehicle tasks, we used a method by Kujala and Mäkelä [17]. The same method has been previously applied to study visual distraction potential of audio-visual route guidance [15] and different text entry methods [14]. This method utilizes visual occlusion technique, originally introduced by Senders, Kristofferson, Dietrich and Ward [28]. Visual occlusion refers to a condition where the driver's vision is occasionally occluded and the duration of the self-selected occlusion is measured. This is later referred as occlusion distance or OD. In this context, visual occlusion is used to measure the distance that is driven during the occluded period, not time. This enables the driver to freely control the driving speed during the measurement of the visual demands of driving.

The testing method is based on an experiment where 97 drivers' occlusion distances on simulated highway and suburban roads were measured [18]. These occlusion distances were mapped on the test routes and used during the distraction testing: the highway routes for participant sample validation and the suburban roads for the actual distraction testing. The participant sample validation by using their occlusion distances driven during an occlusion trial ensure that the driver sample includes both, "short-glancers" and "long-glancers". This validation is an important part of the testing method since previous studies have indicated that drivers have individual off-road glance duration tendencies [2, 16] and these individual differences in glance durations could affect the results of the distraction testing [3, 23].

During the distraction testing, the in-vehicle glance distances are measured. An in-vehicle glance refers to a glance that is directed to an in-vehicle device. Thus, an in-vehicle glance distance refers to a distance in meters that is driven during the in-vehicle glance. These in-vehicle glances can be categorized as green or red glances based on the original 97 drivers' occlusion data [18].

The categorization of the in-vehicle glances is based on the distance driven during a glance from a particular route point where the glance begins. A green glance refers to an in-vehicle glance length that is at or below the baseline data's median occlusion distance for the route point and therefore can be considered as an appropriate glance. The verification threshold for green glances has in previous studies [14, 15] been set to 68 % (min) of all the in-vehicle glances made during the task. To pass the verification criterion, the task should have 68 % or more of green glances. The verification criterion is based on the median percentage of the occlusion distances of the 97 drivers in the study of Kujala, Mäkelä, Kotilainen and Tokkonen [18].

A red glance refers to an in-vehicle glance length that exceeds the 85th percentile of the original 97-driver sample's occlusion distance on the route point. Red glances can thus be considered as inappropriately long in-vehicle glances in relation to the visual demand of the given driving situation. At these occasions, the in-vehicle task has caught the driver's attention for longer time than what the majority of the 97 drivers would have preferred to drive without vision on that route point. The verification threshold for red glances has been set to 6 % (max) [14, 15] of all the in-vehicle glances made during the task. If the task's red glances exceed 6 %, the task fails the verification criterion. The verification criterion is based on the 85th percentile of the occlusion distances of the 97-driver sample in the study of Kujala et al. [18].

3.1 Design of the study

The experimental design for the distraction testing was a within-subjects 2 x 3. The independent variables were the in-vehicle device: a helmet-mounted head-up display or a smartphone, and the in-vehicle task type: navigation, song search or phone call. For studying the gaze concentration effects of the Nuviz HUD (RQ3), the design was a within-subjects 2 x 1 (baseline driving versus route-following with the HUD). The design for NASA-TLX [8] was a within-subjects 4 x 1 the trial as an independent variable (baseline, Nuviz HUD tasks, smartphone tasks, occlusion).

3.2 Participants

The selection of the participants followed the NHTSA [25] recommendations regarding the driver sample for testing distraction of in-vehicle electronic devices, as precisely as possible. Convenience sampling was used to recruit participants via the University of Jyväskylä's mailing lists and connecting local motorcycle clubs. Altogether there were 24 participants (17 males and 7 females). The age of the participants varied from 19 to 72 years (M = 38.7; SD = 14.3). Six participants were 18 to 24 years old, seven were 25 to 39 years old, seven were 40 to 54 years old and four of the participants were older than 55 years. The driving experience varied from 1.5 years to 54 years (M = 20.6; SD = 15.0) and the driven kilometers per year from 6 000 to 40 000 (M = 16 542; SD = 10 283).

3.3 Apparatus

The experiments took place in a driving simulator laboratory of the University of Jyväskylä. A car simulator was used to conduct the experiment although this study is about testing a device that is designed for motorcycle drivers. The driving simulator (see Figure 1) is a medium-fidelity simulator with a motion platform (CKAS Mechatronics 2-DOF). The simulator has automatic transmission, force-feedback steering wheel and pedals (Logitech G27), and the seat is longitudinally adjustable. Eepsoft's (www.eepsoft.fi) professional driving simulator software was used for simulating the driving and saving driving log data at 10 Hz.

The simulator has three 40" LED screens (Samsung, 95.6 cm x 57.4 cm) with resolution of 1440 x 900 per screen. The screens display the driving scene as well as the rear-view mirror and side mirrors. We used a separate 7" tablet (Lenovo TB3-730X) above the steering wheel to display a speedometer to make the position of the meter resemble the meter position in a motorcycle. The tablet received the speed data in near real-time from the simulator software via a Wi-Fi network and the MockGeoFix Android application and displayed the speed to the participants with the Speedometer application available in PlayStore.

During the distraction testing, we used Samsung Galaxy A3 smartphone (4.5", Android 6.0.1), Nuviz head-up display that was mounted to a motorcycle helmet and a controller that was attached to the left side of a steering wheel (see Figure 1). The controller was positioned so that it could be used with the left-hand thumb without taking hands off the steering wheel. The controller is intended to be used in a similar manner when attached to a handlebar of a motorcycle. The smartphone was placed in a holder next to a steering wheel (see Figure 1). A laptop was used to mirror the Nuviz HUD image to ensure that the experimenter saw the same HUD view as the participants.

The Nuviz Android application was running in the same Samsung smartphone, with which the smartphone tasks were conducted. The application communicated with the Nuviz head-up display via a Bluetooth connection. The user interface of the Nuviz HUD can be seen in Figure 2. The functionalities shown in the upper and lower corners at the left side of the views could be selected by the right-hand buttons of the controller. The view (i.e., application or menu position) could be switched up or down by the central scroll button in the controller.

We recorded the eye movements with Ergoneers' Dikablis Essential 50 Hz head-mounted eye-tracking system and synchronized the driving simulator data with the eye-tracking data using LAN bridge.

For the occlusion trial, the steering wheel was outfitted with two levers behind the wheel that reveal the driving scene for 500 milliseconds per pull. If a lever is pulled repeatedly, the driving scene is visible all the time. The time of 500 milliseconds is based on the pioneering research on the occlusion method of Senders et al. [28].

During the trials, we used four predefined routes that simulate actual Finnish suburban roads in the Helsinki metropolitan area. All the routes used during the distraction testing were suburban roads without traffic. The same roads were used also in the study of Kujala and Mäkelä [17]. For the occlusion trial, we used a predefined highway route without any traffic. The driven route was same as for the baseline sample (N = 97) [18]. There were three speed limits: 60 kilometers, 80 kilometers and 120 kilometers per hour. The speed limit changed exactly at the same point for each participant.

3.4 Procedure

Demographic data was collected before the experiment via email. After arrival, the participants signed an informed consent form. At first, each participant adjusted the simulator's seat as close to the steering wheel as possible. This was done to make the HUD image appear above the road environment displayed on the middle screen. The distance between the seat and the steering wheel varied from 32 centimeters to 56 centimeters, mean being 45 centimeters.

Figure 1: Experimental setup and the position of the devices. The orange arrow points to the steering wheel-mounted controller and the green arrow points to the smartphone.

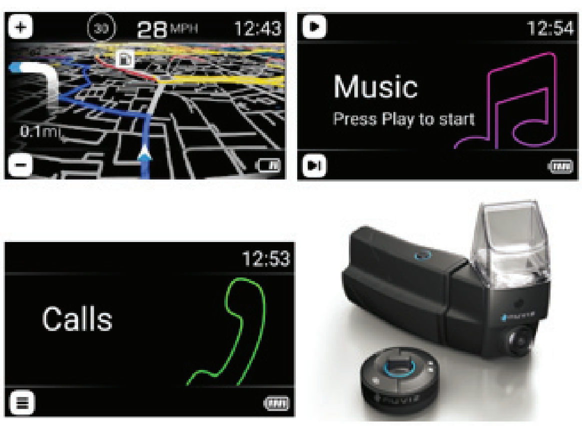

Figure 2: On top: Nuviz HUD user interfaces (navigation and music). On bottom: Nuviz HUD user interface (calls) and Nuviz HUD device with the controller. NB. During the distraction testing the controller did not have any labelling on it.

After adjusting the seat, the participants practiced driving with the simulator as long as they wanted. The practice driving scene was an artificial city environment with other road users. The average practice time was 4.19 minutes. When they started to be familiar with the simulator, they started to practice for the occlusion trial. The purpose of this practice was to get the participants familiar with driving occasionally without vision but still safely. The practice took place in the same artificial city environment with other traffic. The occlusion practice time was on average 6.67 minutes.

The first trial after the two practices was an occlusion trial. The occlusion trial is an important part of the participant sample validation in the testing method. In the occlusion trial all the screens were blank by default and by pulling either of the steering wheel's levers, participants were able to see the driving scene for 500 milliseconds per pull. During the trial, each participant was instructed to obey the traffic rules, to drive safely – but at the same time to try to drive without any visual information of the road (vision occluded) as long as they can.

An extra movie ticket was promised to those six participants who were able to drive the longest distances vision occluded but still accurately. This was done in order to get the participants to focus on the driving task but still trying to maximize the period when they drive without vision. After the trial, each participant filled out the NASA Task Load Index (TLX) questionnaire [8].

After the occlusion trial, the distraction testing started. At this point, the motorcycle helmet with the eye-tracker was put on and the eye-tracker was adjusted and calibrated.

After these preparations, the participants received general driving instructions: to prioritize driving task, to obey the traffic rules and to drive safely. They were also advised that the speed limit is 50 kilometers per hour but they were reminded that they may adjust the speed if needed.

All participants conducted three tasks with the smartphone and with the Nuviz HUD as well as a baseline drive where the task was to follow the verbal navigation instructions provided by the simulator software. The order of the tasks and the routes were counterbalanced in order to avoid learning effects. The baseline drive was always driven between the in-vehicle task trials. The visual demands of each used route were as similar as possible and there were no other road users or red traffic lights in the driving scenarios.

Each participant practiced to conduct the tasks with similar mock-up tasks before the actual task. The smartphone tasks were: 1) to follow the driving instructions to a destination, provided by Google Maps for a pre-defined route, 2) to find a target song from a list of unordered music and to start playing the song using Samsung's native music player (3 songs), and 3) to find a contact information from a list of contacts and to make a call using Samsung's native Contacts application and its call feature (3 calls). Neither of the latter tasks required typing, only scrolling the lists and selecting the correct song or contact information and to tap the call function.

The Nuviz HUD tasks were: 1) to follow the driving instructions to a destination provided by the Nuviz's user interface utilizing a predefined route on HERE maps installed on the Samsung phone, 2) to find a target song from a list of unordered music and to start playing the song using Nuviz's user interface that activated the Samsung's native music player (3 songs), and 3) to find a contact information from a list of contacts and to make a call using Nuviz's user interface that utilized the Samsung's contact information and call feature (3 calls). All the Nuviz tasks were conducted using the physical steering wheel controller that controlled the view in the head-up display. In other words, no typing was required, only scrolling the lists with the central scroll button and making selections by clicking an appropriate button in the corners of the controller.

After each task the participants filled out a reduced NASA-TLX questionnaire without weighting [8]. Finally, each participant was rewarded with a movie ticket.

3.5 Analyses

The main dependent variables for the distraction testing (RQ1) were the percentages of green and red in-vehicle glances and for the analysis of experienced task demands (RQ2) the total NASA-TLX score for each trial. We also report the total number of in-vehicle glances, as well as the total and mean glance duration, and the percentage of over-2-second in-vehicle glances, to provide comparable data with the NHTSA recommended verification criteria [25].

For the analysis of gaze concentration effects of the HUD (RQ3), we compared the standard deviation of the pupil's x-coordinate in eye camera pixels (i.e., horizontal gaze activity) as provided by the Ergoneers' D-Lab software (version 2.5), between the baseline driving and the navigation task with the Nuviz HUD.

The in-vehicle glance lengths were scored following the definition by SAE-J2396 [31]. However, the gaze transition time back to the driving scene was added to a glance duration to provide a 'full' off-road glance length. For the smartphone tasks, glances to the smartphone were counted as in-vehicle glances. For the Nuviz HUD tasks, glances to the HUD and the controller were counted as in-vehicle glances. During the testing, the in-vehicle glances were scored in real-time automatically with a script that recognized pupil's x and y coordinates provided by the Dikablis eye-tracking system. The coordinates were synchronized with the driving simulator's location data. All the automatically scored glances were manually reviewed from synchronized videos (25 fps) using Noldus XY software and all the inaccuracies were corrected frame-by-frame.

To ensure that our driver sample is compatible with the original driver sample [18], the range of the occlusion distances as well as median distances were measured. This was done is order to make sure that the use of the baseline occlusion data is appropriate and that there is no overrepresentation either in "short-glancers" or "long-glancers". Medians were chosen over means because of the non-Gaussian OD distributions. For controlling the effects of accelerations and decelerations in the beginning of the trial, in the junctions and in the end of the trial, only occlusion distances that were driven over 72 kilometers per hour (20 m/s) were included in the data.

Since the distributions of the green and red glances were also non-Gaussian, one-sample sign test was used to test the equality of the green and red glance percentages' medians to the verification thresholds (min 68 % green and max 6 % red). The differences between the Nuviz and smartphone tasks were analyzed with Wilcoxon signed-rank test. The non-parametric Wilcoxon signed-rank test was used also to compare the horizontal gaze activity between the baseline driving and the Nuviz HUD navigation task. For multiple pairwise comparisons, Bonferroni corrections were applied. Differences between the NASA-TLX scores were analyzed with Wilcoxon signed-rank test because most of the distributions were non-Gaussian. Where applicable, Cohen's d is reported as a measure of effect size.

4 Results

4.1 Occlusion distances

Due to technical problems in one trial, N of occlusion distances is 23. The occlusion distance varied from 9.0 to 37.4 meters (range 28.4 m), median being 20.7 meters. The equivalent range and median of the baseline data [18] are 3.2 - 41.9 meters (38.7 m) and 13.7 meters. Out of interest, there was a strong inverse correlation between occlusion distance and age: $r = -.50$.

4.2 Mean number of in-vehicle glances

Table 1 indicates that there were enough glances per task type for meaningful statistical testing. The mean numbers of glances for the song search and call tasks in Table 1 can be multiplied by three to get the total number of glances analyzed in the distraction testing.

Device	Navigation	Song search	Call
Nuviz HUD	74.50 (36.90)	19.24 (8.50)	8.78 (5.70)
Samsung smartphone	46.96 (21.47)	14.74 (11.65)	12.99 (7.97)

Table 1: Mean number of in-vehicle glances per task type. Standard deviation in parentheses. The song search and call tasks are averaged over three tasks.

4.3 Green in-vehicle glances

The verification threshold for green glances was set to 68 % (min) [14, 15]. According to one-sample sign test, the Nuviz HUD task passed the verification criterion for green glances (see Figure 3), the percentage being 68.64 ($p = .376$). The smartphone tasks did not pass the criterion since the median was 42.26 % ($p < .001$). Wilcoxon signed-rank test indicated that there is a significant difference in the green glance percentages between the Nuviz HUD and smartphone tasks: $Z = 3.49$, $p < .001$. The effect is large ($d = 0.90$).

Wilcoxon signed-rank test indicated that at task-level only statistically significant difference in green glance percentages between the devices, in favor of Nuviz HUD, was with the song search task: $Z = 3.71$, $p < .001$, $d = 1.39$ (large effect), with a Bonferroni-corrected alpha level of .017 (see Table 2). For the other two tasks, the pairwise differences were not statistically significant after Bonferroni-correction: navigation $Z = 2.29$, $p = .022$, $d = 0.75$; call $Z = 2.23$, $p = .026$, $d = 0.53$.

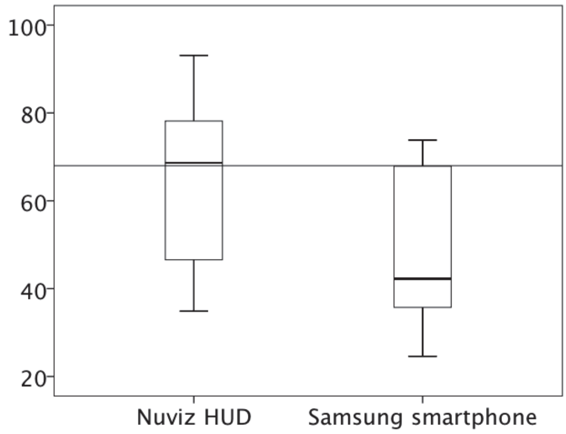

Figure 3: Percentage of green in-vehicle glances, verification threshold at 68 %.

Device	Navigation	Song search	Call
Nuviz HUD	56.6	73.6	68.3
Samsung smartphone	32.8	36.2	54.0

Table 2: Percentage of green in-vehicle glances (median) per input method and task type.

4.4 Red in-vehicle glances

The verification threshold of red glances was set to 6 % (max) [14, 15]. According to one-sample sign test, both – Nuviz HUD and smartphone – tasks passed the set verification criterion (see Figure 4). Nuviz's median red glance percentage was 3.41 ($p = .511$) and smartphone's 6.20 ($p = .162$). The smartphone's red glance percentage does not differ significantly from the threshold (6 %) and therefore passed the test.

According to Wilcoxon signed-rank test, there is also a significant difference in the red glance percentages between the Nuviz HUD and smartphone tasks: $Z = 2.74$, $p = .006$. The effect is of medium size ($d = 0.62$).

Wilcoxon signed-rank test indicated that there was a significant difference between the devices (see Table 3), in favor of Nuviz HUD, in the song search task ($Z = 2.66$, $p = .008$, $d = 0.82$ [large effect]) and in the navigation task ($Z = 2.40$, $p = .016$, $d = 0.57$ [medium effect]), with the Bonferroni-corrected alpha level of .017. No difference was found in the call task ($Z = .501$, $p = .616$).

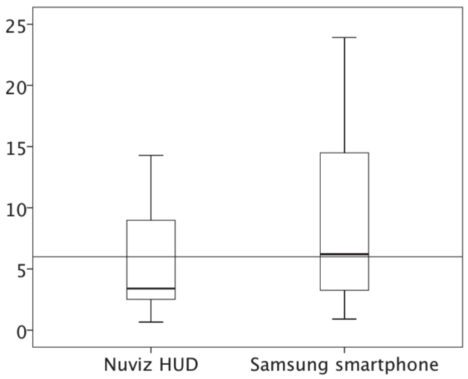

Figure 4: Percentage of red in-vehicle glances, verification threshold at 6 %.

Device	Navigation	Song search	Call
Nuviz HUD	3.8	2.0	4.1
Samsung smartphone	9.1	8.0	3.2

Table 3: Percentage of red in-vehicle glances (median) per input method and task type.

4.5 Durations of in-vehicle glances (NHTSA, 2013)

For enabling comparison between studies, also the NHTSA [25] recommended metrics are reported in Tables 4-6.

	Navigation		Song search		Call	
Device	M (SD)	85th %ile	M (SD)	85th %ile	M (SD)	85th %ile
Nuviz HUD	56.29 (17.69)	74.88	14.30 (3.21)	17.06	6.88 (2.32)	10.23
Samsung smartphone	50.56 (20.67)	77.65	16.22 (9.21)	26.12	11.41 (4.02)	15.87

Table 4: Total duration of in-vehicle glances (s). Standard deviation in parentheses. The song search and call tasks are averaged over three tasks.

	Navigation		Song search		Call	
Device	Md	85th %ile	Md	85t %ile	Md	85th %ile
Nuviz HUD	0.0	3.5	0.0	2.1	0.0	5.8
Samsung smartphone	0.0	5.9	8.6	25.3	4.7	12.8

Table 5: Percentage of over-2-second in-vehicle glances (median). Percentages calculated for three tasks per task type for the song search and call tasks.

	Navigation		Song search		Call	
Device	M (SD)	85th %ile	M (SD)	85th %ile	M (SD)	85th %ile
Nuviz HUD	0.86 (0.27)	1.13	0.81 (0.23)	1.09	0.87 (0.25)	1.14
Samsung smartphone	1.13 (0.22)	1.33	1.23 (0.35)	1.60	1.00 (0.35)	1.38

Table 6: Mean in-vehicle glance durations (s). Means calculated for three tasks per task type for the song search and call tasks.

4.6 Experienced task workload - NASA-TLX

According to Wilcoxon signed-rank test, all the differences between the trials were significant with α = .008, except the difference between occlusion trial and smartphone tasks (p = .158, see Figure 5): baseline vs. Nuviz HUD tasks, Z = 3.42, p = .001, d = 0.73; baseline vs. occlusion, Z = 4.17, p < .001, d = 2.20; baseline vs. smartphone tasks, Z = 4.26, p < .001, d = 1.78; Nuviz HUD tasks vs. occlusion, Z = 3.93, p < .001, d = 1.36; and Nuviz HUD tasks vs. smartphone tasks, Z = 3.33, p = .001, d = 1.02. All the effect sizes are large, except between baseline driving and Nuviz HUD tasks the effect size is medium.

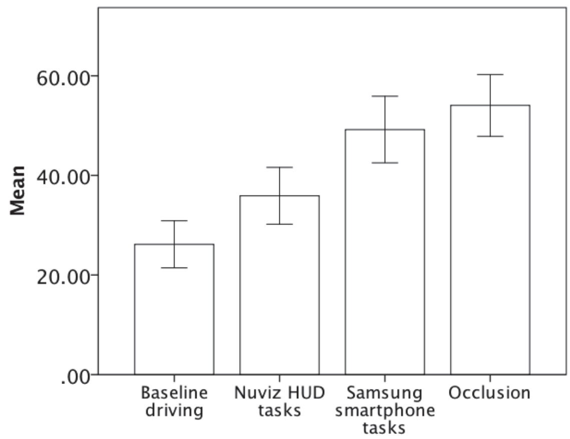

Figure 5: Experienced task workload measured with NASA-TLX. Maximum is 100. Error bars: 95 % CI.

4.7 Horizontal gaze activity

The horizontal gaze activity was measured with standard deviation of pupil's x-coordinate in eye camera pixels (see Figure 6). According to Wilcoxon signed-rank test, there was no significant difference in horizontal gaze activity between the baseline driving and the Nuviz HUD navigation task (p = .349).

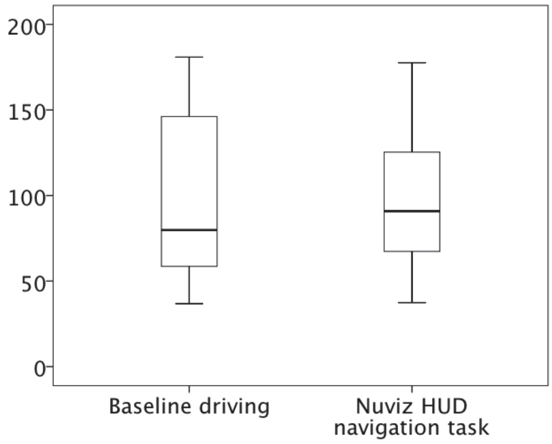

Figure 6: Standard deviation of pupil's x-coordinate in eye camera pixels.

5 Discussion

This was the first study comparing visual distraction effects between tasks conducted with a helmet-mounted HUD for motorcyclists and with a smartphone. The distraction potential of these two devices were assessed with red and green in-vehicle glances as defined by Kujala and Mäkelä [17]. The verification threshold for green glances was set to 68 % (min) and for red glances maximum to 6 % (max) [14, 15]. Overall, the Nuviz HUD tasks passed the set verification criterion for green glances, the percentage of green glances being 68.6. The smartphone tasks did not pass this criterion with the percentage of 42.3. When compared per task type, the song search task conducted with Nuviz HUD had significantly higher green glance percentages than the similar task with a smartphone (here, higher is better).

Both – Nuviz HUD and smartphone tasks – passed the set verification criterion for the inappropriately long red glances. The overall percentage for the Nuviz HUD tasks was 3.4 and for the smartphone tasks 6.2. The smartphone tasks passed the criterion as well because the difference between the percentage and the verification threshold is not significant. However, the difference in red glance percentages between the Nuviz HUD and smartphone tasks is significant. When compared per task type, the song search and navigation tasks had significantly lower red glance percentages when conducted with the Nuviz HUD.

Based on these findings, the studied Nuviz HUD tasks seem to have lower distraction potential than the tasks with the same goals conducted with an Android smartphone (RQ1). Compared to the smartphone tasks, the Nuviz HUD tasks increased the percentage of green in-vehicle glances by 62 % and decreased the percentage of red in-vehicle glances by 45 %. Also, there was no overrepresentation of "short-glancers" or "long-glancers" in the participants' occlusion distance distribution. Based on that, the sample can be considered comparable with the baseline 97-driver sample [18] and the green and red glance metrics as reliable.

The experienced task workload was reported highest during the occlusion trial and second highest during the smartphone tasks. No significant difference between these two trials were found. The baseline drive was experienced as the least demanding and the Nuviz HUD tasks were experienced as second least demanding. There was a significant difference between baseline driving and Nuviz HUD tasks as well as between Nuviz HUD tasks and smartphone tasks. The Nuviz HUD tasks were experienced less demanding than the smartphone tasks (RQ2).

We found no difference in horizontal gaze activity between baseline driving and the navigation task conducted with the Nuviz HUD. It can be interpreted that using HUD for navigation while driving did not cause gaze concentration in this study (RQ3). The possible gaze concentration effects of HUDs should be further studied also with other types of in-vehicle tasks than navigation.

Due to confounding factors, we cannot pinpoint the exact design factors explaining the advantage of the tasks with the Nuviz HUD over the similar tasks with a smartphone. However, based on earlier research, we assume that the HUD enabled use of peripheral vision to maintain better situational awareness of the demands of the driving environment during the HUD tasks compared to the head-down phone tasks. In addition, the easy-to-learn haptic design of the HUD controller could be used without vision. We noticed that glances directed to the steering wheel-mounted HUD controller were very rare, from a few to none. This is a positive sign towards the haptic design of the controller.

Previous studies (e.g., [35, 36]) have found similar results concerning HUDs operated with physical controllers and touchscreen HDDs. However, the tested Nuviz HUD differs from those since it is designed to be used while driving a motorcycle and the HUD is helmet-mounted, ensuring that the HUD view is visible for the driver in all head positions. This is not the case with windshield HUDs [35, 36]. Windshield HUDs [35, 36] cannot be used in a motorcycle context since motorcycles are missing car-like windshields and that is why a different HUD solution is needed.

As an example of UI design differences at the software level, both music search tasks required scrolling a list of music to find the target song(s). With the Nuviz HUD the participant could scroll the list song-by-song by a single press of a physical button in the steering wheel controller. With the smartphone, the participant had to point the touchscreen and scroll the music player menu by means of kinetic scrolling, which has been found to be one of the most visually distracting activities with touchscreen in-vehicle devices [19, 20]. In addition, the participant did not have to look down, far away from the driving environment, to see the selected song in the HUD.

The navigation, song search, and call tasks had equivalent or even higher mean number and total duration of in-vehicle glances conducted with the Nuviz HUD than with the smartphone (Table 1). Despite that, the Nuviz HUD tasks had higher green glance percentages and lower red glance percentages than the tasks conducted with the smartphone. This finding suggests, that these metrics of visual demand may suit poorly for measuring visual distraction. This is the case, in particular, if visual distraction is operationalized as a calibration failure between the situational visual demands of driving and the off-road glance length. Similarly, Smith, Gabbard and Conley [29] found out that HUD affected more negatively the NHTSA glance metrics [25] than HDD. Depending on the user interface, the task length may not be as critical factor for appropriate timing of the in-vehicle glances and distraction than other task features. The finding stresses the importance of in-vehicle user interface and task design to mitigate visual distraction and the importance of using proper metrics suitable for a user interface design in distraction testing. The green and red glance percentages observed in this study are well in line with previous findings [14, 15]. This gives credibility to the used distraction testing method as similar task designs seem to produce similar results.

Nonetheless, there are some limitations concerning this study. The car simulator used in this study cannot simulate driving with a motorcycle. In addition, in this study the Nuviz controller was attached to the steering wheel (see Figure 1). When driving a motorcycle, the controller would be attached to a handlebar. However, the study was designed to enable comparative analysis of the visual distraction effects of device used while driving (a simulated car) in a controlled environment. Motorcycling can be argued to be more demanding than driving a car [24], and thus, the absolute distraction effects may be even larger while riding a motorcycle than what measured here. Naturally this applies to both smartphone tasks and HUD tasks. The measured visual distraction effects of the in-vehicle tasks cannot be generalized to provide estimates of the absolute distraction effects while driving a motorcycle, but we argue that the observed relative effects between the devices and tasks are reliable. In fact, the generalization of any driving performance or glance data measured in a simulator to real conditions has to be done with caution.

Road surface roughness was absent because the simulator's motion platform has only two degrees of freedom. This factor could favor the Nuviz HUD with the thumb-controller even more in real traffic conditions. All the in-vehicle tasks in this study, also with the smartphone, were relatively easy due to low number of task steps. This was due to the fairly limited Nuviz HUD functionalities at the time of testing. With more complex in-vehicle tasks the distraction effects of both smartphone and HUD tasks could be worse. One should also keep in mind the usability-distraction paradox: the overall distraction effects in a driver population may be increased by safer and easier-to-use in-vehicle user interfaces, if these increase the frequency of use of these devices on the roads.

6 Conclusion

This was the first research comparing visual distraction effects of a HUD designed for motorcyclists to distraction effects of smartphone usage. The distraction effects were evaluated with a novel method that classifies in-vehicle glances to appropriate and inappropriate glances dependent on the situational driving demands. Compared to the smartphone tasks, the Nuviz HUD tasks increased the percentage of acceptable in-vehicle glances by 62 % and decreased the percentage of inappropriately long in-vehicle glances by 45 %. Based on the results, the tested HUD tasks seem to be safer for motorcyclists than similar tasks with a smartphone while driving.

The tasks conducted with the HUD where also reported less demanding than the tasks conducted with the smartphone. The use of the HUD for navigation guidance did not cause gaze concentration effects compared to baseline driving. However, these effects are something to be studied more carefully in the future with other types of in-vehicle tasks.

The study had some limitations since a car simulator was used instead of a motorcycle simulator. On the other hand, driving a motorcycle is more complex task than driving a car and that is why the distraction effects can be even larger than what reported while riding a motorcycle.

ACKNOWLEDGMENTS

We thank Sara Norja for conducting the experiments. The study was partly funded by Nuviz Inc. The funder had no role in data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

[1] Stewart A. Birrell and Mark S. Young. 2011. The Impact of Smart Driving Aids on Driving Performance and Driver Distraction. *Transportation Research Part F: Traffic Psychology and Behaviour*, 14(6), 484–493. DOI: https://doi.org/10.1016/j.trf.2011.08.004

[2] Robert Broström, Peter Bengtsson and Mikael Ljung Aust. 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. DOI: http://dx.doi.org/10.1016/j.trf.2015.10.017

[3] Robert Broström, Mikael Ljung Aust, Linnea Wahlberg and Laban Källgren. 2013. What Drives Off-Road Glance Durations During Multitasking – Capacity, Practice or Strategy. In *Proceedings of the 3rd International Conference on Driver Distraction and Inattention*.

[4] David E. Crundall and Geoffrey Underwood. 1998. Effects of Experience and Processing Demands on Visual Information Acquisition in Drivers. *Ergonomics*, 41, 448-458. DOI: https://doi.org/10.1080/001401398186937

[5] Department of Environment, Transport and the Regions (DETR) Report. 2000. Tomorrow's Roads—Safer for Everyone: The Government's Road Safety Strategy and Casualty Reduction Targets for 2010, HMSO.

[6] Gregory M. Fitch, Susan A. Soccolich, Feng Guo, Julie McClafferty, Youjia Fang, Rebecca L. Olson, Miguel A. Perez, Richard J. Hanowski, Jonathan M. Hankey and Thomas A. Dingus. 2013. The Impact of Hand-Held and Hands-Free Cell Phone Use on Driving Performance and Safety-Critical Event Risk. *Report No. COT HS 811 757. National Highway Traffic Safety Administration, U.S. Department of Transportation*.

[7] James P. Foley, Richard Young, Londa Angell and Joshua E. Domeyer. 2013. Towards Operationalizing Driver Distraction. In *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*. DOI: 10.17077/drivingassessment.1467

[8] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in psychology*, 52, 139-183. DOI: https://doi.org/10.1016/S0166-4115(08)62386-9

[9] Jibo He, William Choi, Jason S. McCarley, Barbara S. Chaparro and Chun Wang. 2015. Texting while Driving Using Google Glass™: Promising but not Distraction-Free. *Accident Analysis & Prevention*, 81, 218-229. DOI: https://doi.org/10.1016/j.aap.2015.03.033

[10] Jibo He, Jason S. McCarley, Kirsten Crager, Murtuza Jadliwala, Lesheng Hua and Sheng Huang. 2018. Does Wearable Device Bring Distraction Closer to Drivers? Comparing Smartphones and Google Glass. *Applied Ergonomics*, 70, 156-166. DOI: https://doi.org/10.1016/j.apergo.2018.02.022

[11] Katja Kircher and Christer Ahlstrom. 2017. Minimum Required Attention: A Human-Centered Approach to Driver Inattention. *Human Factors*, 59(3), 471-484.

[12] Sheila G. Klauer, Thomas A. Dingus, Vicki L. Neale, Jeremy D. Sudweeks and D. J. Ramsey. 2006. The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data. *National Highway Traffic Safety Administration, U.S. Department of Transportation*.

[13] Sheila G. Klauer, Feng Guo, Bruce G. Simons-Morton, Marie Claude Ouimet, Suzanne E. Lee and Thomas A. Dingus. 2014. Distracted Driving and Risk of Road Crashes among Novice and Experienced Drivers. *New England Journal of Medicine*, 370(1), 54-59.

[14] Tuomo Kujala and Hilkka Grahn. 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 ACM,* 1-10. DOI: https://doi.org/10.1145/3122986.3122987

[15] Tuomo Kujala, Hilkka Grahn, Jakke Mäkelä and Annegret Lasch. 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, ACM,* 169-176. DOI: https://doi.org/10.1145/3003715.3005421

[16] Tuomo Kujala, Annegret Lasch and Jakke Mäkelä. 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*, 1-8. DO: https://doi.org/10.1145/2667317.2667322

[17] Tuomo Kujala and Jakke Mäkelä. 2015. Development of a Testing Environment and a Verification Procedure for In-Vehicle Tasks with Dynamic Driving Scenarios. In *Proceedings of the 4th International Conference on Driver Distraction and Inattention*.

[18] Tuomo Kujala, Jakke Mäkelä, Ilkka Kotilainen and Timo Tokkonen, 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. DOI: https://doi.org/10.1177/0018720815595901

[19] Tuomo Kujala, Johanna Silvennoinen and Annegret Lasch. 2013. Visual-Manual In-Car Tasks Decomposed: Text Entry and Kinetic Scrolling as The Main Sources of Visual Distraction. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 82-89. DOI: https://doi.org/10.1145/2516540.2516562

[20] Annegret Lasch and Tuomo Kujala. 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*, 41-48.

[21] Felix Lauber, Claudius Böttcher and Andreas Butz. 2014. You've Got the Look: Visualizing Infotainment Shortcuts in Head-Mounted Displays. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications,* 1-8. DOI: https://doi.org/10.1145/2667317.2667408

[22] Ja Young Lee, Madeleine C. Gibson and John D. Lee. 2016. Error Recovery in Multitasking While Driving. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5104-5113. DOI: https://doi.org/10.1145/2858036.2858238

[23] Ja Young Lee and John D. Lee. 2017. Multi-level Analysis of Distracted Drivers' Glances: Enhancing the Robustness of the NHTSA Acceptance Criteria. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 61(1), 1919-1923.

[24] Fred L. Mannering and Lawrence L. Grodsky. 1995. Statistical Analysis of Motorcyclists' Perceived Accident Risk. *Accident Analysis & Prevention*, 27(1), 21-31.

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

[26] Rebeca L. Olson, Richard J. Hanowski, Jeffrey S. Hickman and Joseph Bocanegra. 2009. Driver Distraction in Commercial Vehicle Operations. *Report No. FMCSARRR-09-042. Federal Motor Carrier Safety Administration, U.S. Department of Transportation*.

[27] Sysavanh Phommachanh, Masao Ichikawa, Shinji Nakahara, Mayfong Mayxay and Akio Kimura. 2017. Student Motorcyclists' Mobile Phone Use while Driving in Vientiane, Laos. *International Journal of Injury Control and Safety Promotion*, 24(2), 245-250. DOI: https://doi.org/10.1080/17457300.2016.1166141

[28] John W. Senders, A. B. Kristofferson, W. H. Levison, C. W. Dietrich and J. L. Ward. 1967. The Attentional Demand of Automobile Driving. *Highway Research Record*, 195.

[29] Missie Smith, Joseph L. Gabbard and Christian Conley. 2016. Head-Up vs. Head-Down Displays: Examining Traditional Methods of Display Assessment while Driving. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 185-192. DOI: https://doi.org/10.1145/3003715.3005419

[30] Missie Smith, Jillian Streeter, Gary Burnett and Joseph L. Gabbard. 2015. Visual Search Tasks: The Effects of Head-Up Displays on Driving and Task Performance. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications,* 80-87. DOI: https://doi.org/10.1145/2799250.2799291

[31] 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.

[32] Long T. Truong, Chris De Gruyter and Hang T. T. Nguyen. 2017. Calling, Texting, and Searching for Information while Riding a Motorcycle: A Study of University Students in Vietnam. *Traffic Injury Prevention*, 18(6), 593-598. DOI: https://doi.org/10.1080/15389588.2017.1283490

[33] Trent Victor, Marco Dozza, Jonas Bärgman, Christian-Nils Boda, Johan Engström, Carol Flannagan, John D. Lee, and Gustav Markkula. *Analysis of Naturalistic Driving Study Data: Safer Glances, Driver Inattention, and Crash Risk*. Washington, DC: Transportation Research Board.

[34] Trent Victor, Joanne L. Harbluk and Johan A. Engström. 2005. Sensitivity of Eye-Movement Measures to In-Vehicle Task Difficulty. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8(2), 167-190. DOI: https://doi.org/10.1016/j.trf.2005.04.014

[35] Gabriella Villalobos-Zúñiga, Tuomo Kujala and Antti Oulasvirta. 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,* 177-184. DOI: https://doi.org/10.1145/3003715.3005453

[36] Garrett Weinberg, Bret Harsham and Zeljko Medenica. 2011. Evaluating the Usability of a Head-Up Display for Selection from Choice Lists in Cars. In *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications, ACM,* 39-46.

VI

IMPACTS OF TOUCH SCREEN SIZE, USER INTERFACE DESIGN, AND SUBTASK BOUNDARIES ON IN-CAR TASK'S VISUAL DEMAND AND DRIVER DISTRACTION

by

Hilkka Grahn & Tuomo Kujala, 2020

International Journal of Human-Computer Studies, 142, 102467

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

International Journal of Human-Computer Studies 142 (2020) 102467

Contents lists available at ScienceDirect

International Journal of Human-Computer Studies

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

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

Hilkka Grahn*, Tuomo Kujala

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

A R T I C L E I N F O

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

A B S T R A C T

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

1. INTRODUCTION

Over the past decade, the effects of smartphone usage on traffic safety have been in the focus of research all over the world. Several studies have explored the detrimental impacts of using smartphone while driving with different methods. For instance, there have been many naturalistic (e.g., Guo et al., 2010; Tivesten and Dozza, 2015) and simulator studies (e.g., Choudhary and Velaga, 2017; He et al., 2015a; Rumschlag et al., 2015) as well as surveys (e.g., Bayer and Campbell, 2012; Gauld et al., 2017) and meta-analyses (e.g., Caird et al., 2014; Oviedo-Trespalacios et al., 2016) investigating drivers' smartphone use and its effects on driver performance and safety. As a general finding, these studies have strengthened the association between smartphone usage and drivers' visual distraction.

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

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

That said, one design factor that could mitigate drivers' visual distraction is to utilize speech-to-text function in order to decrease the visual-manual demands of an in-vehicle system. Speech-to-text technology recognizes driver's speech and converts it into commands that the system can understand. Various studies have assessed the efficacy of speech-to-text function (or voice recognition) to mitigate driver distraction compared to manual text entry (e.g., Beckers et al., 2017; He et al., 2015b, 2014; Tsimhoni et al., 2004). Typically, manual text entries are nowadays conducted with touch screen keyboards – which

* Corresponding author.
E-mail addresses: hilkka.grahn@jyu.fi (H. Grahn), tuomo.kujala@jyu.fi (T. Kujala).

https://doi.org/10.1016/j.ijhcs.2020.102467
Received 27 May 2019; Received in revised form 6 March 2020; Accepted 13 May 2020
Available online 24 May 2020
1071-5819/

are visually highly demanding and causing distraction for drivers (e.g., Crandall and Chaparro, 2012; Kujala et al., 2013; Kujala and Grahn, 2017; McKeever et al., 2013; Reimer et al., 2014b). Another design factor that could mitigate drivers' visual distraction is to utilize read-aloud function which is a technology that reads selected text aloud. However, there is little published glance duration data on read-aloud function in the driving context. According to Owens et al. (2011), read-aloud function does not cause longer glances away from the road compared to baseline driving. Conversely, other studies that are not based on glance duration metrics have found that read-aloud functions may not be distraction-free either (Jamson et al., 2004; Lee et al., 2001). Actually, Reimer and Mehler (2013) and Reimer et al. (2014a) have pointed out that both speech-to-text and read-aloud functions are rarely completely free of visual or manual interaction and can therefore be visually distracting. There are also some task types, such as checking all nearby gas stations from a navigation system, that may be inefficient to be conducted using verbal communications only as the read-aloud function would orally list all the options.

Previous research has also established, in general, that screen size affects efficiency when conducting different tasks (Hancock et al., 2015; Raptis et al., 2013). Gaffar and Kouchak (2017) studied in automotive context drivers' reaction times when selecting target icon on either 7" or 10" screen. They concluded that there was no difference in reaction times between those two relatively large screens. No glance durations were measured in their study. Hence, to our best knowledge, there are no existing well-controlled studies in automotive context about the effects of screen size on glance durations, comparing for instance a smartphone screen versus a tablet screen. Similarly, previous studies have not extensively dealt with screen orientation's effects on visual distraction in detail.

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

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

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

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

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

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

2. GENERAL METHOD

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

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

In the visual distraction potential testing, the operationalization of visual distraction is based on the data collected by Kujala et al. (2016b), where 97 drivers' preferred occlusion distances on simulated highway and suburban roads were measured while the drivers were only focusing on safe driving. Afterwards the measured occlusion distances were mapped on the same test routes. This means that each 1×1 meter route point in the map holds information on occlusion distances that were driven in that particular route point in the experiment. When the same test routes are used in the visual distraction potential testing, it enables the categorization of in-car glances (i.e., glances that are directed to the in-car device) to appropriate or inappropriate. The categorization is based on the distance driven during the in-car glance from a particular route point where the glance begins. An inappropriate or *red in-car glance* refers to an in-car glance length that exceeds the 85th percentile of the 97-driver sample's occlusion distance on the route point. Red glances can thus be considered as inappropriately long in-car glances in relation to the visual demand of the given driving situation and which can be considered as visual distraction.

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

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

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

3. Experiment 1

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

3.1. Method

3.1.1. Experimental design

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

3.1.2. Participants

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

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

3.1.3. Apparatus

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

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

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

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

3.1.4. Procedure

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

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

drive when vision is occasionally occluded. The environment for the occlusion trial was the same artificial city with other road users as in the previous practice but the starting point was different. The average practice time was four and half minutes.

First task after the practices was the occlusion trial for the validation of the test sample. During the occlusion trial the driving simulator's screens were blank by default and the driving scenery could be revealed for 500 milliseconds (as in Senders et al., 1967) by pressing the levers

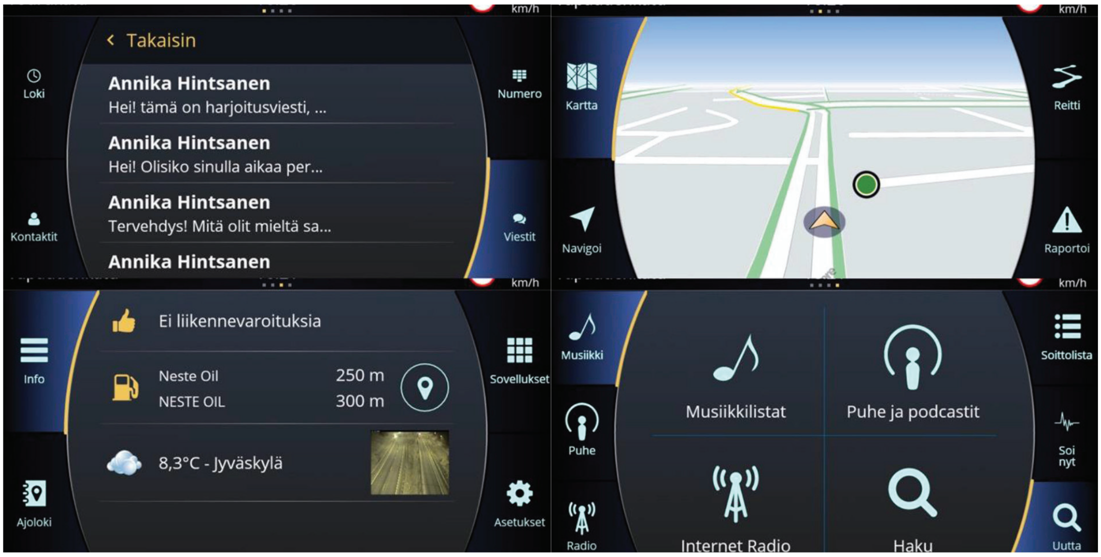

Figure 2. Views of the automotive-targeted application (Carrio). Left upper corner: email reading view, right upper corner: navigation view, left lower corner: weather condition view, and right lower: corner song searching view. The navigation view and the weather condition view were only used in the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

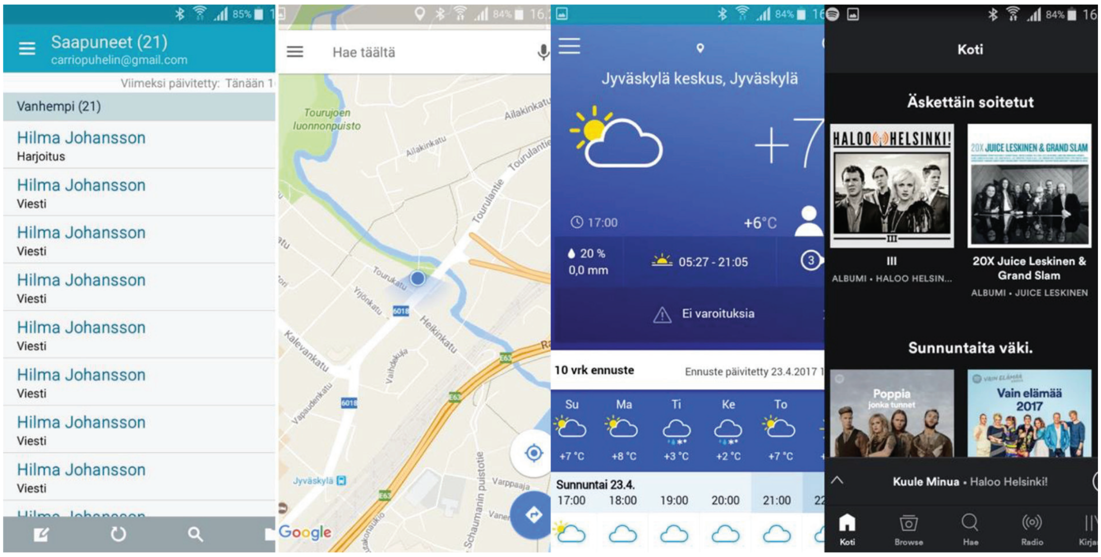

Figure 3. Views of regular smartphone applications (Android). From left: email reading view, navigation view, weather condition view, and song searching view. The navigation view and the weather condition view were only used in the view switching task.

that were attached to the steering wheel. The two-lane highway route without traffic was the same that was used to gather the baseline data of Kujala et al. (2016b). Before the trial, each participant received instructions to obey the traffic regulations, to drive safely and accurately but still to try to drive without visual information (i.e., vision occluded) as long as possible. Those six participants who could drive the longest median distances without visual information and still accurately, were promised a movie ticket as an extra reward. The reward was promised in order to get participants to concentrate on the driving task but still trying to maximize the occlusion distance to their preference. The speed limits varied from 60 to 80 to 120 kilometers per hour depending on the highway section. Every change in the speed limit was told to each participant at the same point of the route. However, sections that were driven 60 km/h were junctions from a highway to another highway and were not included in the final data. After the trial, NASA-TLX questionnaire (Hart and Staveland, 1988) was filled out in order to measure task workload.

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

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

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

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

3.1.5. Data preparation

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

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

3.2. Results and discussion

3.2.1. Occlusion distances

To validate the driver sample, the distribution of the occlusion distances was compared to the original occlusion distance distribution of the 97-driver sample (Kujala et al., 2016b) where the occlusion distances varied from 3.21 meters to 41.88 meters ($Mdn = 13.67$). In

Table 1
Tasks in Experiment 1.

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

this experiment, the occlusion distances varied from 6.35 meters to 35.82 meters with a median of 17.37 meters. According to Levene's test, the variance of the occlusion distance distribution does not differ significantly from the original OD distribution of Kujala et al. (2016b): (F (1,116) = .645, p = .424).

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

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

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

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

3.2.4. Discussion

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

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

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

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

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

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

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

Table 2
Task procedures in Experiment 1.

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

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

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

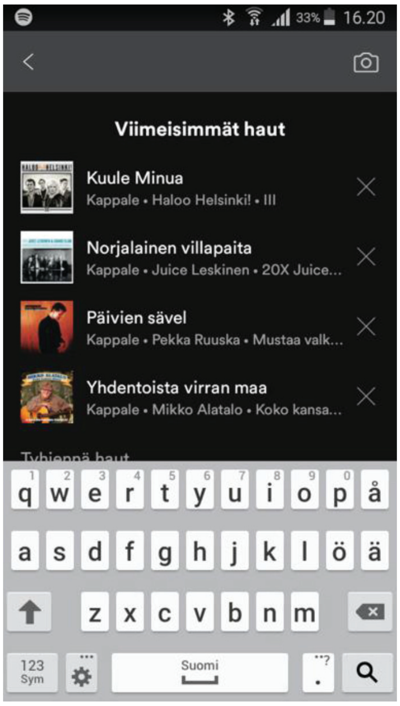

Figure 4. The qwerty touch screen keyboard of the Android smartphone.

smartphone had 2.5” smaller screen and was used in portrait mode – this arrangement could have also favored Carrio.

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

4. Experiment 2

Our aim in Experiment 2 was to examine if the observed lower distraction potential of the Carrio tasks in Experiment 1 was due to the larger tablet screen or its landscape orientation compared to the smaller-sized smartphone in portrait mode. To be able to compare the effects of these factors between the experiments, participants conducted in this experiment the same email reading and song searching tasks as in Experiment 1. However, this time both Carrio and regular smartphone applications were running on 4.5” smartphone in landscape mode. The view-switching task was omitted since it was found relatively easy and low-distracting in Experiment 1 with both, Carrio and Android. Instead, participants had an extra task to reply to four emails to get data from a text entry task with manual typing compared to a speech-to-text function. Based on previous research (e.g., Crandall and Chaparro, 2012; McKeever et al., 2013; Reimer et al., 2014b), the manual text entry task was assumed to be the most visually distracting due to task structure (preferred typing of word per glance) and manual keyboard input requiring high accuracy.

4.1. Method

4.1.1. Experimental design

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

4.1.2. Participants

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

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

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

4.1.3. Apparatus

The exactly same driving simulator (see Figure 1), other equipment (excluding the tablet), routes (see Figure 5), and statistical tools were used in this experiment. A software update to the commercial Carrio application between the experiments enabled us to study speech-to-text function in the text entry task. However, an additional smartphone had to be used for this task to keep the same older version of Carrio for the email reading and song searching tasks in the same phone as in Experiment 1. Changing the versions of Carrio during the experiment was evaluated to be too risky since it could have caused technical difficulties during the experiment. In addition, this would have extended the duration of experiments. Since the same version of Samsung Galaxy A3 did not exist in the market anymore, the additional phone was Samsung Galaxy A3 (2017) with Android 7.0 operating system. It has 4,7” screen which is 0.2” larger than in Samsung Galaxy A3 used in Experiment 1. In each task both smartphones were used in landscape mode in order to be able to control the effects of the screen orientation.

4.1.4. Procedure

Each participant went through exact same practices as in Experiment 1. In the artificial city scenario, the average practice time was five minutes, and in the occlusion drive the average practice time

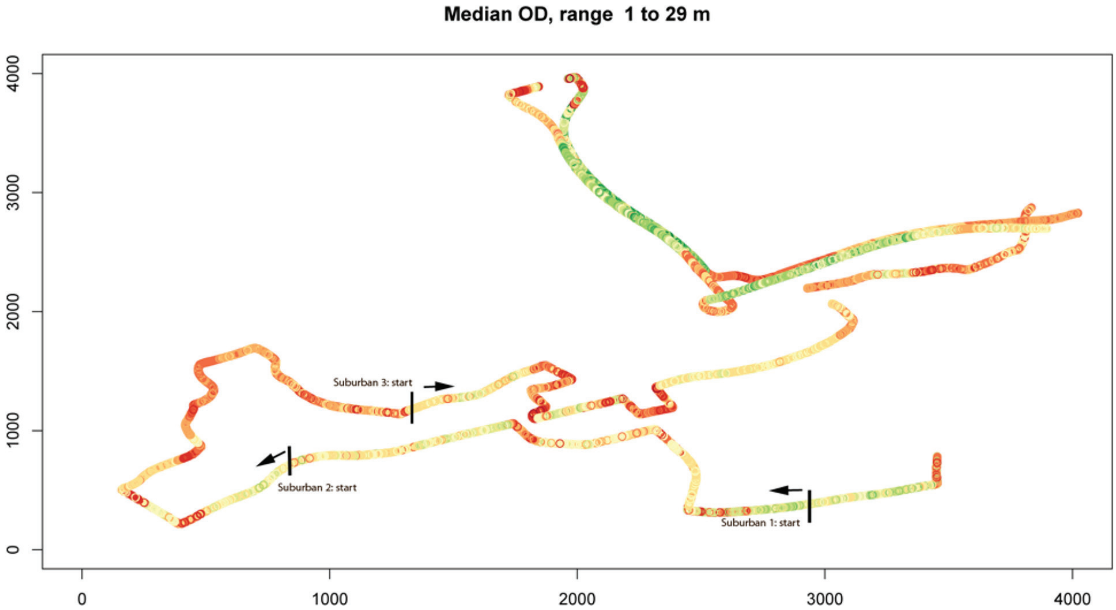

Figure 5. The pre-defined routes for the experiments. Color indicates the visual demand of the route point as occlusion distance: the demand increases from green to yellow to orange to red. The routes are same as in Kujala and Mäkelä (2015) and Kujala et al. (2016b).

was four minutes. After the practices, the occlusion trial was conducted exactly the same as in Experiment 1. The general instructions, routes, experimenter, counterbalancing, and practicing the mock tasks were the same as in Experiment 1.

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

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

4.1.5. Data preparation

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

4.2. Results and discussion

4.2.1. Occlusion distances

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

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

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

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

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

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

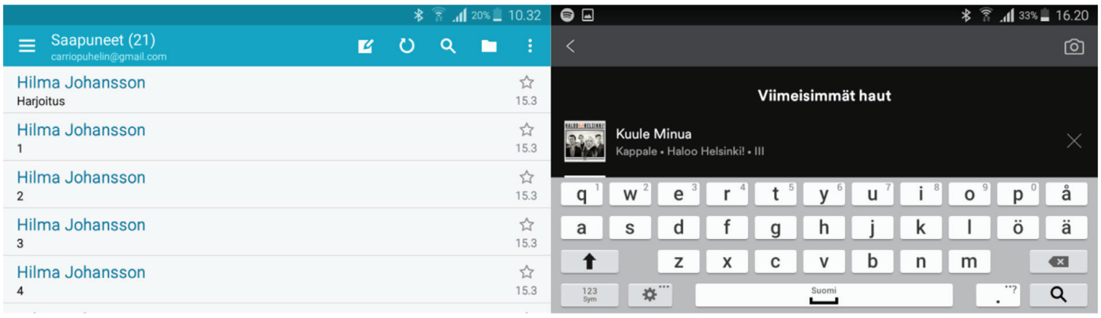

Figure 7. Views of regular smartphone applications (Android). From left: email reading view and song searching view.

4.2.4. Discussion

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

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

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

5. Multilevel modeling and analyses of design factors

5.1. Model 1

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

5.1.1. Screen size, screen orientation, and application

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

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

The equation for the first model is:

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

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

In the model, the grand mean of the in-car glance duration is 932 milliseconds for Carrio on the larger 7" screen. Compared to the Carrio application, the use of regular smartphone applications (Android) increase the in-car glance duration by 279 milliseconds and when the size of the screen decreases from 7" to 4.5", the duration of the in-car glance increases by 39 milliseconds. The model indicates also that one-meter increase in occlusion distance – which can be interpreted as inverse of visual demand of the driving situation – increases the duration of the in-car glance by 12 milliseconds. This means that there is a 120-milliseconds increase in in-car glance duration by 10 meter increase in occlusion distance. In other words, when the driving scenario was less

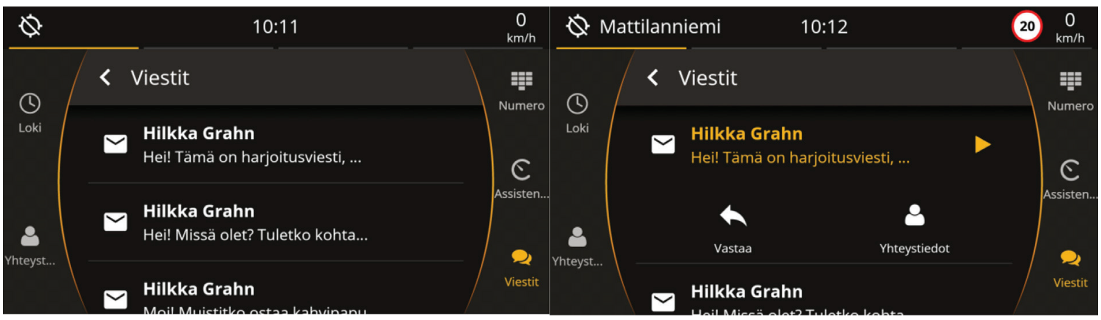

Figure 8. Views of the automotive-targeted application (Carrio). From left: email replying view and view when the email is being listened to.

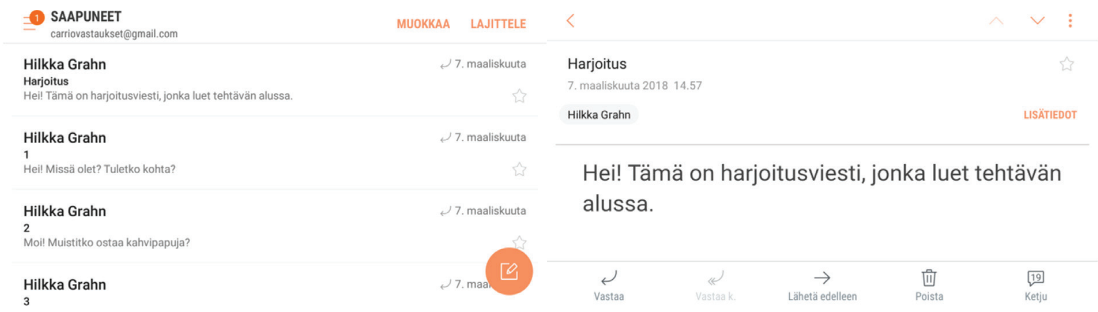

Figure 9. Views of regular smartphone application (Android). From left: email view and view when the email is being replied to.

visually demanding (e.g., no junctions ahead), participants were able to glance the in-car device longer.

5.2. Model 2

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

5.2.1. Task features

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

The equation for the second model is:

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

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

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

5.2.2. Tasks grouped by visual demand

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

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

All the emails in the reading tasks started with a short greeting, other than that, the emails contained on average four sentences of meaningful information. Based on the mean number of glances, participants made on average 4.34 glances per email in Experiment 1 and 4.28 glances in Experiment 2. This indicates that participants read one sentence per glance, on average. To complete one task in the email replying task conducted with Android, participants had to read an email (1–2 meaningful short sentences) and type an answer (2–3 words, 16.75 characters on average). In addition, four button presses were required. Based on the mean number of glances, to read and type one email required on average 12.85 glances. The average should be close to 21.75 glances if the participants had typed a single character per glance. Here, end of a sentence or finishing typing a word can be assumed to be the self-selected subtask boundary that offered a natural break point to participants. In general, subtask boundaries are used while multitasking to switch attention in natural break points to reduce cognitive load (e.g., Janssen et al., 2012; Payne et al., 2007) and our findings are consistent with this idea.

Table 5
Tasks in Experiment 2.

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

Table 6
Task procedures in Experiment 2.

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

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

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

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

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

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

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

* The factor above is compared to factor that gets the value of zero.
** Occlusion distance: inverse of visual demand of the driving situation.

As in the group of visually high demanding tasks, the intermediate group also had one deviation among the red in-car glance percentages. Carrio's email reading task conducted with the 4.5” screen had higher red in-car glance percentage than some tasks in the visually high demanding group. The observed errors made during the selection of the next email (see Figure 8) could have an effect on these red glance percentages. That is, some participants selected same emails twice since they were confused which email was most recently selected. Alternatively, the presence of email's first line of text (see Figure 8) could have affected both red in-car glance percentages and mean number of in-car glances: participants may have read the presented line of text instead of just listening to the email. Again, hypothetically, the errors may have caused participants to glance the in-car device at a road point where they would have not glanced without this cognitive distraction (Lee et al., 2016, 2007). However, unfortunately we do not have access to error data in these tasks. Otherwise, the range of the red in-car glance percentages is well in line with the visual demand grouping.

Finally, there was only one task in the group of visually low demanding tasks: view-switching task with Carrio. This task was relatively easy since the order of the views seemed to be easily learnable and this

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

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

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

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

*The factor above is compared to factor that gets the value of zero.
**Occlusion distance: inverse of visual demand of the driving situation.

enabled drivers to switch views with a simple gesture while looking at the road ahead. This task had the lowest mean number of in-car glances (see Table 3). This is a similar finding as in previous studies that found simple gestures for scrolling pages one-by-one to be the most visually least demanding and distracting when compared to button presses or kinetic scrolling (e.g., Kujala, 2013; Lasch and Kujala, 2012).

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

6. GENERAL DISCUSSION

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

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

Since these results from Experiment 1 and Experiment 2 individually did not clarify the effects of the different design factors, we constructed two multilevel models based on the data from both experiments. With multilevel models we investigated how screen size, screen orientation, application, and task affect in-car glance durations in the tested tasks. Multilevel modeling enabled us to control the effects of visual demand of the driving scenario and individual differences on in-car glance durations. Together with the accompanying data from the experiments, these models indicated that, in general, Carrio tasks had lower visual demand and visual distraction potential compared to the tasks conducted with the regular smartphone applications (Research Questions 1 and 2).

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

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

Further, based on the overlaps in the confidence intervals in the Model 2, we identified three task groups (see Table 11): visually high demanding, visually intermediately demanding, and visually low demanding tasks (Research Question 2) which all have their own common

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

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

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

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

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

6.1. Limitations and further research

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

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

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

7. CONCLUSIONS

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

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

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

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

CRediT authorship contribution statement

Hilkka Grahn: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing - original draft. **Tuomo Kujala:** Conceptualization, Methodology, Funding acquisition, Project administration, Writing - review & editing, Supervision.

Declaration of Competing Interest

None.

Acknowledgements

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

References

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

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

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

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

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

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

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

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

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

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

Grahn, H., Kujala, T., 2018. Visual distraction effects between in-vehicle tasks with a smartphone and a motorcycle helmet-mounted head-up display. In: ACM International Conference Proceeding Series, pp. 153–162. https://doi.org/10.1145/3275116.3275134.

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

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

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

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

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

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

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

Hox, J., 1998. Multilevel modeling: When and why. In: Balderjahn, I., Mathar, R., Schader, M. (Eds.), Classification, Data Analysis, and Data Highways. Studies in

Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, pp. 147–154. https://doi.org/10.1007/978-3-642-72087-1_17.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Lee, Y.-C., Lee, J.D., Ng Boyle, L., 2007. Visual attention in driving: The effects of cognitive load and visual disruption. Hum. Factors J. Hum. Factors Ergon. Soc. 49, 721–733 https://doi.org/10.1518/001872007 × 215791.

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

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

National Highway Traffic Safety Administration, 2013. Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices, DocketNo. NHTSA-2010-0053. https://doi.org/NHTSA-2010-0053.

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

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

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

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

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

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

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

Richter, T., 2006. What is wrong with ANOVA and multiple regression? Analyzing sentence reading times with hierarchical linear models. Discourse Process 41, 221–250. https://doi.org/10.1207/s15326950dp4103_1.

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

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

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

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

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

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

Wierwille, W.W., 1993. In: Gale, A.G., Brown, I.D., Haslegrave, C.M., Kruysse, H.W., Taylor, S.P. (Eds.), Eds. Vision in Vehicles IV.

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

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

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